INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

ISOTOPIC AND GEOCHEMICAL INVESTIGATIONS OF PRECAMBRIAN CONTINENTAL CRUST IN THE TORNGAT OROGEN, NORTHEASTERN CANADA: CONSTRAINTS ON THE MECHANISMS OF PRECAMBRIAN CRUST FORMATION AND ON THE EARLY PROTEROZOIC ASSEMBLY OF NORTHEASTERN LAURENTIA

by

LISA MARIANNE CAMPBELL

B.S., University of California, Santa Cruz, 1986

M.A., University of California, Santa Cruz, 1990

A thesis submitted to the

Faculty of the Graduate School,

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Geological Sciences

1997

This thesis entitled:

ISOTOPIC AND GEOCHEMICAL INVESTIGATIONS OF PRECAMBRIAN CONTINENTAL CRUST IN THE TORNGAT OROGEN, NORTHEASTERN CANADA: CONSTRAINTS ON THE MECHANISMS OF PRECAMBRIAN CRUST FORMATION AND ON THE EARLY PROTEROZOIC ASSEMBLY OF NORTHEASTERN LAURENTIA

by

LISA MARIANNE CAMPBELL

has been approved for the Department of Geological Sciences

Dr. G. Lang Farmer

Dr. Charles R. Stern

Date 1/9/98

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline

Campbell, Lisa Marianne (Ph.D., Geological Sciences)

ISOTOPIC AND GEOCHEMICAL INVESTIGATIONS OF PRECAMBRIAN CONTINENTAL CRUST IN THE TORNGAT OROGEN, NORTHEASTERN CANADA: CONSTRAINTS ON THE MECHANISMS OF PRECAMBRIAN CRUST FORMATION AND ON THE EARLY PROTEROZOIC ASSEMBLY OF NORTHEASTERN LAURENTIA

Thesis directed by Professor G. Lang Farmer

ABSTRACT

The Torngat Orogen (TO) in Northern Labrador, Canada, formed during Early Proterozoic continental collision between the Archean Nain and Rae provinces. The TO preserves Early Proterozoic and Archean rocks which offer an excellent opportunity to study crust-mantle evolution and collisional tectonics during the Early Precambrian. This dissertation presents the results of isotopic and geochemical investigations on Archean and Early Proterozoic metaigneous rocks from this region. The studies were designed to: (a) document the petrogenesis of these rocks; (b) determine the lateral extent of Early Proterozoic metaigneous rocks and underlying Archean basement gneisses; and (c) evaluate spatial variations in the isotopic compositions of Early Proterozoic metaigneous rocks with respect to current models for subduction zone geometry.

The study has shown that (a) the Burwell domain (BD) is underlain by Archean gneisses which differ laterally in isotopic and geochemical composition and are "Nain-like" in the northeast, and "Rae-like" in the southwest. This suggests that the BD is not allochthonous with respect to adjacent Archean Provinces and contains elements of both; (b) Early Proterozoic metaigneous rocks occurring within the BD and Nain Province are geochemically similar to Phanerozoic arc magmas, however spatial variations in their $\varepsilon Nd_{(i)}$ values do not increase inland (eastward), as would be expected for the proposed eastward dipping subduction models. It is suggested here that the spatial distrubution of $\varepsilon Nd_{(i)}$ reflect emplacement of the arc magmas into Archean crust which was underplated by a variably thick column of basalt, possibly emplaced

during a period of incipient rifting prior to subduction. Although there is currently no direct evidence for this, there are several pieces of indirect evidence that rifting may have preceeded arc magmatism in the eastern BD

In previous models for the assembly of NE Laurentia, the BD represented an independent crustal terrane which was accreted to the Nain Provice during Proterozoic collision. The results from this study suggest that this "foreign" terrane did not exist, and hence that revisions to the models are necessary. This research further confirms that modern-style plate tectonic processes controlled Early Proterozoic crustal amalgamation in this segment of Northeastern Laurentia.

DEDICATION

To my Father, Ballard Crooker Campbell (October, 1915 - December, 1996). He taught me the important things. Among these, to respect life and to challenge my own.

ACKNOWLEDGEMENTS

The work presented in this thesis was possible due to the help, cooperation and encouragement of a large number of people. I would first like to thank my thesis advisor, Lang Farmer, for all of his help during the course of my work, particularly for the discussions which helped me to develop my ideas, not only about the topics of the study, but also on the larger, global scale, and to see the implications of my work. I would like to thank Emily and Phillip Verplank, as well as Lang, for the excellent training I recieved in developing the multitude of skills required for clean lab chemistry and mass spectrometry.

I would also warmly like to thank Dick Wardle (Newfoundland Dept. of Mines and Energy) and Martin Van Kranendonk (Geological Survey of Canada), who provided the opportunity of a life-time.... to study the rocks (and avoid the wildlife), of northern-most Labrador, Canada through participation in their Northern Torngat project. The field experiences were invaluable in terms of learning to map and decipher complex, high-grade gneiss terrains. Discussions with them and with the rest of the field party were enlightening, motivating and fun. Field mapping as a part of the team was an excellent experience, both scientifically and personally... the friendships which developed are priceless.

Professor David Bridgwater (Geological Museum, Copenhagen, Denmark), shared with me the insights into Precambrian geology which he developed over a lifetime. In addition he enthusiastically shared with me his knowledge of the rocks not only in northern Labrador, but in the Nagssugtoqidian in West Greenland, the Lewisian in NW Scotland and on the White Sea in NW Russia. He is a mentor and a friend.

Discussions with many people (in addition to Lang Farmer and others mentioned above), have helped me develop and refine my ideas and I wish to thank Dave Scott, Jim Connelly, Jane Selverstone, Chuck Stern, Lon Abbott, Flemming Mengel and Jan Dunphy. Fellow graduate students and staff at the Geology Dept., University of Colorado, Boulder have offered help, guidance and camaraderie over the years. I would particularly like to thank Chuck Patterson, Lisa Dierauf and Rich McDonald for their constant friendship, companionship and support, and for keeping me feed during protracted work attacks.

Lastly I thank my husband, Flemming Mengel for tireless support, love, friendship and advice. He has somehow, miraculously managed to stay by my side, despite living 6,000 miles away.

TABLE OF CONTENTS

SIGNATURE PAGE ii
ABSTRACT iii
DEDICATION v
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS vii
LIST OF TABLES xii
LIST OF FIGURES xiii

CHAPTER 1

Introduction

1.1 Introduction and Background 1
1.1.1 Introduction 1
1.1.2 Problems in Precambrian Geology 5
1.2 Torngat Orogen - Presentation of Study Area 9
1.2.1 Introduction 9
1.2.2 Outstanding questions related to development of the Torngat Orogen: status at the start of this project 10
1.2.3 Overview of previous geological studies in the Torngat Orogen 11
1.2.4 Thesis goals 15
1.2.5 Field component of the project: Logistics 16
1.3 Methodology 17
1.3.1 Introduction 17
1.3.2 Nd isotope system 17
1.3.3 Sr isotope system 20
1.3.4 Pb isotope system 21
1.4 Structure of the Thesis 22
1.4.1 Introduction 22
1.4.2 Chapter 2 23
1.4.3 Chapter 3 23
1.4.4 Chapter 4 24
1.5 References 25

CHAPTER 2

Archean Crustal Blocks in Northern Labrador: Isotopic and Geochemical Constraints on Archean Crustal Development Prior to Torngat Orogenesis

2.1 Introduction 30

2.2 Geological Overview 33

2.2.1 The Nain Province 34

2.2.2 The Rae Province 37

2.2.3 The Burwell Domain 38

2.3 Scope of the Study 39

2.4 Description of the samples 40

2.4.1 Nain Province 40

2.4.1.1 Field Description 40

2.4.1.2 Petrography 40

2.4.2 Rae Province 41

2.4.2.1 Field Description 41

2.4.2.2 Petrography 41

2.4.3 Archean gneisses in the Burwell Domain 42

2.4.3.1 Field Description 42

2.4.3.2 Petrography 42

2.5 Presentation of the Data 43

2.5.1 Major and Trace Element Geochemistry 43

2.5.1.1 Nain Province 43

2.5.1.2 Rae Province 58

2.5.1.3 Archean gneisses in the Burwell Domain 61

2.5.2 Isotope Geochemistry 62

2.5.2.1 Nain Province 62

2.5.2.2 Rae Province 68

2.5.2.3 Archean gneisses in the Burwell Domain 69

2.6 Discussion 69

2.6.1 Tonalitic gneisses in northern Labrador: Comparisons with high-Al TTDs 69

2.6.2 Significance of Eu-anomalies 71

2.6.3 Geochemical effects of granulite facies metamorphism 71

2.6.4 The compositie nature of the Nain Province 73

2.6.4.1 Origin of the anorthosites 76

2.6.5 Rae Province-comparison with the Nain Province 76
2.6.5.1 Age and origin of granitic rocks in the Rae Province 77
2.6.5.2 Origin of amphibolites in the Rae Province 78
2.6.6 Burwell Domain- links to the Rae and Nain Province? 78
2.6.7 Summary 79
2.7 Conclusions 81
2.8 References 82

CHAPTER 3

Isotopic And Geochemical Constraints On Petrogenesis Of Early Proterozoic Igneous Rocks In The Northern Torngat Orogen, Labrador: Implications For Tectonic Assemby Of The Torngat Orogen.

3.1 Introduction 87
3.2 Regional Geology of the Torngat Orogen 90
3.2.1 The Northern Torngat Orogen 91
3.2.2 Current models for development of Torngat Orogen 92
3.3 Scope of this study 93
3.4 Description of samples 95
3.4.1 Group 1 95
3.4.1.1 Field description 95
3.4.1.2 Petrography 95
3.4.2 Group 2 96
3.4.2.1 Field description 96
3.4.2.2 Petrography 96
3.4.3 Group 3 97
3.4.3.1 Field description 97
3.4.3.2 Petrography 97
3.5 Presentation of Data 97
3.5.1 Major and Trace Element Geochemistry 97
3.5.1.1 Geochemistry of Group 1 samples 103
3.5.1.2 Geochemistry of Group 2 samples 110
3.5.1.3 Geochemistry of Group 3 samples 113
3.5.2 Isotope Geochemistry 115
3.5.2.1 Group 1 115
3.5.2.2 Group 2 115
3.5.2.3 Group 3 122

3.6 Discussion 123
3.6.1 Geochemical Effects of Granulite facies metamorphism 123
3.6.2 Origin of Group 1 rocks 125
3.6.2.1 Tholeiites 125
3.6.2.2 Calc-alkaline rocks 127
3.6.3 Origin of Group 2 rocks 128
3.6.4 Origin of Group 3 rocks 130
3.6.5 Origin of the Early Proterozoic magmas 131
3.6.5.1 Modeling of Group 1-Fractional Crystallization 132
3.6.5.2 AFC Modelling of Groups 1-3 134
3.6.6 Significance of Geographic Distribution of $\varepsilon ND_{(1895)}$ 138
3.7 Conclusions 142
3.8 References 146

CHAPTER 4

Models for the growth of Laurentia: revisions based on investigations in the Torngat Orogen, northern Labrador, Canada

4.1 Introduction 150
4.2 Early Proterozoic assembly and growth of Laurentia: the model of Hoffman (1988) 151
4.2.1 Cape Smith Belt 153
4.2.2 New Quebec Orogen 153
4.2.3 Foxe Fold Belt 153
4.2.4 Torngat Orogen: formed by Nain-Rae collision 155
4.3 Early Proterozoic assembly and growth of Northeastern Laurentia: the model of Hoffman (1990) 155
4.4 Early Proterozoic assembly and growth of Northeastern Laurentia: the model of Van Kranendonk et al.(1993) - differences with previous models 156
4.5 Models for development of the Torngat Orogen: different pre-collisional geometries 160
4.5.1 Westward subduction prior to collision - data from southern TO 160
4.5.2 Eastward subduction prior to collision - new data and reinterpretation of existing data 161

4.5.3 Double subduction or flip in subduction-direction? 163

4.6 Modifications to existing tectonic models: implications from geochemical and isotopic investigations 165

4.6.1 Origin of Early Proterozoic meta-igneous rocks: a magmatic arc built on the Nain margin - summary of data and interpretations 165

4.6.2 The Burwell Domain: an Early Proterozoic magmatic complex built on Archean Nain Province 166

4.6.3 Anorthosites and amphibolites: evidence for pre-collisional rifting? 167

4.6.4 Tasiuyak gneiss: Passive margin sediments or an accretionary wedge?..... 168

4.7 Early Proterozoic assembly of NE Laurentia: constraints from isotopic and geochemical studies in the northern Torngat Orogen - summary and conclusions 171

4.8 References 174

BIBLIOGRAPHY 178

APPENDIX 1

A.1 Sample preparation 189

A.2 Major and trace element analyses 189

A.3 Rb-Sr and Sm-Nd Isotopic analyses 190

A.3.1 Element Separation Chemistry 190

A.3.2 Mass Spectrometric Methods 191

A.4 Pb Isotopic analyses 193

LIST OF TABLES

CHAPTER 2

Table 1 Major element analyses of Archean gneisses from the Nain Province, Rae Province and Burwell domain, northern Torngat Orogen, Labrador 44

Table 2 Trace element analyses from Archean gneisses from the Nain Province, Rae Province and Burwell domain, northern Torngat Orogen, Labrador 46

Table 3 Sr, Nd and Pb Isotopic analyses of Archean gneisses from the Nain Province, Rae Province and Burwell domain, northern Torngat Orogen, Labrador 63

CHAPTER 3

Table 1 Major element analyses of Early Proterozoic metaigneous rocks, northern Torngat Orogen, Labrador 98

Table 2 Trace element analyses Early Proterozoic metaigneous rocks, northern Torngat Orogen, Labrador 100

Table 3 Sr, Nd and Pb Isotopic analyses of Early Proterozoic metaigneous rocks, northern Torngat Orogen, Labrador 116

CHAPTER 4

Table 1 Nd Isotopic analyses of Tasiuyak gneiss metasediments, Torngat Orogen, northern Labrador 170

LIST OF FIGURES

CHAPTER 1

Figure 1. Precambrian tectonic elements of Laurentia from Hoffman, 1988. The Baltic shield is shown in a pre-lapetus reconstruction., and Greenland is restored to its pre-rift position relative to North America. Uppercase names are Archean provinces; lowercase names are Proterozoic and Phanerozoic orogens. Abbreviations: BH, Black Hills inlier,; BL, Belcher belt; CH, Cheyenne belt; CL, Cape Smith belt, FR, Fox River belt; GL, Great Lakes tectonic zone; GS, Great Slave Lake shear zone; KL, Killarney magmatic zone; KP, Kapuskasing uplift; KR, Keweenawan rift zone; LW, Lapland-White Sea tectonic zone; MK, Makkovik orogen; MO, Mistassini-Otish basins; MRV, Minnesota River Valley terrane; SG, Sugluk terrane; TH, Thompson belt; TS, Transscandinavian (Smaland-Varmland) magmatic zone; VT, Vetrenny tectonic zone; WR, Winisk River fault. 2

Figure 2. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. Also shown are the location of Figure 3 ("This study") 3

Figure 3. Geological sketch map of the study area in Northern Labrador (for location, see Fig.2) Based on data from Wardle et al. (1993), Van Kranendonk et al. (1994). ASZ and KSZ are theAbloviak and Komaktorvik shear zones, respectively 4

CHAPTER 2

Figure 1. Simplified geological map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements, sample locations and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Early Proterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. 31

Figure 2. Geological sketch map of the study area in Northern Labrador (for location, see Figure 1). Based on data from Wardle et al. (1993), Van Kranendonk et al. (1994). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. 32

Figure 3. Sketch map of Northern Labrador showing the location of major lithotectonic units discussed in the text, including the Archean Saglek and Hopedale blocks of the Nain Province, and the Paleoproterozoic Makkovik subprovince. The area covered in the study by Bardoux et al. (1996) in Ungava Bay is also indicated. Inset map shows the linkages between the Archean Saglek and Hopedale blocks in Labrador and the Archean block in Southern Greenland. This North Atlantic Archean Craton is bordered to the West and North by the Paleoproterozoic Torngat (T) and Nagssugtoqidian (N) orogens, and to the South by the Makkovikian (M) orogen. Modified from Bridgwater and Schiotte, 1990 and Collerson et al., 1982. 35

Figure 4. Compositional classification of Archean rocks from the Torngat Orogen, Northern Labrador. (A) - AFM diagram (Irvine & Baragar, 1971). FeO* = total Fe calculated as FeO. (B) - CIPW normative Anorthite-Albite-Orthoclase diagram (fields from Barker, 1979). (C) - CIPW normative quartz-alkali-feldspar-plagioclase diagram from Le Maitre (1989), after Strekeisen (1976). 49

Figure 5. Selected major elements vs. SiO_2, for Archean rocks from the Torngat Orogen, Northern Labrador. Total Fe is given as Fe_2O_3. 50

Figure 6. Selected major and trace elements vs. SiO_2 for Archean rocks from the Torngat Orogen, Northern Labrador. 51

Figure 7. Selected trace and rare earth elements vs. SiO_2 for Archean rocks from the Torngat Orogen, Northern Labrador. 52

Figure 8. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). (A)+(B)+(C) - Tonalitic gneisses from the Nain Province. Samples distributed between diagrams based on REE patterns. (D) - Gneisses of intermediate compositions from the Nain Province. 53

Figure 9. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). (A) - Amphibolites from the Nain and Rae provinces. (B) - Anorthosites from the Nain Province and Burwell Domain. 55

Figure 10. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). (A) - Tonalitic gneisses from the Rae Province. (B) - Granitic rocks from the Rae Province. (C) - Tonalitic gneisses from the Burwell Domain. Data for island arc calc-alkaline basalts from Wilson (1989)....56

Figure 11. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). (A)+(B)+(C) - Tonalitic rocks from the Nain Province distributed between diagrams according to REE patterns (see Figure 8A, B and C). 57

Figure 12. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). (A) - Tonalitic gneisses of intermediate composition from the Nain Province. (B) - Amphibolites from the Nain and Rae provinces. (C) - Anorthosites from Nain Province and Burwell Domain. Data for continental arc tholeiitic basalt from Ewart (1982), and for island arc tholeiitic basalt from Sun (1980). 59

Figure 13. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). (A) - Tonalitic gneisses from the Rae Province. (B) - Granitic rocks from Rae Province. (C) - Tonalitic rocks from the Burwell Domain. 60

Figure 14. Sm/Nd isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. (A) - εNd vs. age diagram showing Nd isotopic evolution lines for a range of lithologies. Depleted mantle curve from DePaolo (1981). (B) - εNd (M) vs. $^{147}Sm/^{144}Nd$ for a range of lithologies. Reference isochron of 2.97 Ga is based on samples from the Rae Province (see also text). 65

Figure 15. Rb/Sr and Sm/Nd isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. (A) - Rb/Sr isochron diagram for all samples analyzed. Reference isochron of 3044 Ma is based on all data points. (B) - $^{87}Sr/^{86}Sr$(M) vs. εNd (M) diagram for all samples analyzed. 66

Figure 16. Pb isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. (A) - $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ isotope correlation diagram. Reference isochron of 3087 Ma is based on all data. Stacey-Kramers two stage evolution curve (Stacey & Kramers, 1975) is shown for reference. (B) - $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ isotope correlation diagram. The "Bulk Earth Closed System Evolution" curve (k = 4.2) is from Gariépy and Dupré (1991) and is shown for reference. 67

Figure 17. Plot of Rb (ppm) versus K (wt.%) for Archean tonalitic, trondhjemitic and granodioritic gneisses in the northern Torngat Orogen. The diagram also shows the "Granulite Trend" (GT) from Rudnick et al. (1985) and the "Main Trend" of continental igneous rocks (granites through basalts) from Shaw (1968). Horizontal arrows represent the amount and direction of Rb depletion caused by granulite facies metamorphism (from Rudnick et al., 1985). 74

CHAPTER 3

Figure 1. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. Also shown are the location of Figure 2 ("This study") and the area described by Theriault & Ermanovics (1997). 88

Figure 2. Simplified geological map of the study area in northernmost Labrador (modified from Wardle et al., 1993; Van Kranendonk et al., 1994). For location, see Figure 1. ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively (see also Figure 1). Sample locations and distribution of members of Groups 1, 2 and 3 are also shown (see text for description of Groups). Symbols are as follows: samples in Group 1 are shown as diamonds; Group 2 as squares, Group 3 as circles. 89

Figure 3. Geographic distribution of $\varepsilon_{Nd\,(1895)}$ values for Early Proterozoic metaigneous rocks in the Torngat Orogen, Northern Labrador. Basemap identical to Figure 2. Contours based on values of $\varepsilon_{Nd\,(1895)}$, i.e., $\varepsilon_{Nd\,(1895)} > 0$ (Group 1), $-5 < \varepsilon_{Nd\,(1895)} < 0$ (Group 2) and $\varepsilon_{Nd\,(1895)} < -5$ (Group 3). 94

Figure 4. Compositional classification of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. (A) - AFM diagram (Irvine & Baragar, 1971). FeO* = total Fe calculated as FeO. (B) - Anorthite-Albite-Orthoclase diagram (fields from Barker, 1976). (C) - CIPW mesonormative quartz-alkali-feldspar-plagioclase diagram from Le Maitre (1989). 104

Figure 5. Selected major elements vs. SiO_2.for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Fields of ocean island basalts (OIB) and mid-oceanic ridge basalts (MORB) from Wilson (1989). 105

Figure 6. Selected major and trace elements vs. SiO_2.for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. 106

Figure 7. Selected trace and rare earth elements vs. SiO_2.for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. 107

Figure 8. Chondrite normalized REE patterns from Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). (A)+(B) - Calc-alkaline rocks in Group 1 ($\varepsilon_{Nd(1895)} > 0$). (C) - Tholeiitic rocks in Group 1 ($\varepsilon_{Nd(1895)} > 0$). Data from island arc basalt (IAB) and mid-ocean ridge basalt (MORB) from Wilson (1989). 108

Figure 9. N-MORB normalized trace element data (Spider diagram) for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values and sequence of variables from McCulloch and Gamble (1990). **(A)+(B)** - Calc-alkaline samples in Group 1. **(C)** - Tholeiitic samples in Group 1. Data for continental arc basalt from Ewart (1982), data for island arc tholeiitic basalt from Sun (1980), and data for calc-alkaline island arc basalt from Sun (1980). 109

Figure 10. Chondrite normalized REE patterns for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). All data from Group 2 samples ($-5<\varepsilon_{Nd(1895)} < 0$). Samples distributed in (A), (B) and (C) according to shape of REE patterns. 111

Figure 11. N-MORB normalized trace element data (Spider diagram) for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values and sequence of variables from McCulloch and Gamble (1990). All data from Group 2 samples ($-5<\varepsilon_{Nd(1895)} < 0$). Distribution of samples in (A), (B) and (C) based on shapes of REE patterns (see Figure 9). Data for continental arc basalt from Ewart (1982) and data for island arc calc-alkaline island arc basalt from Sun (1980). 112

Figure 12. Trace and rare earth element data from Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - Chondrite normalized REE patterns from Group 3 samples ($\varepsilon_{Nd(1895)} < -5$). Normalizing values from Evensen (1978). Shaded field represent compositional variation among Nain and Rae Province Archean gneisses (data from Chapter 2). **(B)** - N-MORB normalized trace element data from Group 3 samples ($\varepsilon_{Nd(1895)} < -5$). Normalizing values and sequence of elements from McCulloch and Gamble (1990). Data for continental arc basalt from Ewart (1982) and data for island arc calc-alkaline island arc basalt from Sun (1980). 114

Figure 13. Sm/Nd isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen Northern Labrador. **(A)** - $\varepsilon_{Nd\ (1895)}$ vs. $^{147}Sm/^{144}Nd$ diagram for samples from Group 1, 2 and 3. Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). **(B)** - $\varepsilon_{Nd\ (M)}$ vs. $^{147}Sm/^{144}Nd$ for samples from Group 1, 2 and 3. Stippled field includes data from Nain and Rae Province Archean gneisses (Chapter 2). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). 118

Figure 14. Rb/Sr and Sm/Nd isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - $^{87}Rb/^{86}Sr$ vs. $^{87}Sr/^{86}Sr(I)$ isotope correlation diagram for samples from Group 1, 2 and 3. **(B)** - Rb/Sr isochron diagram for samples from Group 1, 2 and 3. Fields of Nain and Rae Province and Burwell Domain tonalites based on data from Chapter 2. Isochrons at 1.87 Ga and 2.14 Ga are shown for reference. 119

Figure 15. Pb isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ isotope correlation diagram showing the distribution of samples from Group 1, 2 and 3. Also shown are fields of Nain and Rae Province and Burwell Domain tonalites based on data from Chapter 2. Stacey-Kramers two stage evolution curve (Stacey & Kramers, 1975) is shown for reference. **(B)** - $^{238}U/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ isotope correlation diagram showing distribution of Group 1 and 2 samples. 120

Figure 16. Pb isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ isotope correlation diagram showing the distribution of samples from Group 1, 2 and 3. Also shown are fields of Nain and Rae Province and Burwell Domain Archean tonalites based on data from Chapter 2. 121

Figure 17. Plot of Rb (ppm) versus K (wt.%) for Early Proterozoic metaigneous rocks from Torngat Orogen. The diagram also shows the "Granulite Trend" (GT) from Rudnick et al. (1985) and the "Main Trend" of continental igneous rocks (granites through basalts) from Shaw (1968). Horizontal arrows represent the amount and direction of Rb depletion caused by granulite facies metamorphism (from Rudnick et al., 1985). 124

Figure 18. Results of fractional crystallization of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Calculations for fractional crystallization based on equations in Rollinson (1993). Nd (ppm) vs. Sr (ppm) showing the result of fractional crystallization using LC92-40 as the primary melt composition. 133

Figure 19. Initial Nd and Sr isotope composition vs. SiO_2 for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - Initial εNd vs. wt% SiO_2. **(B)** - Initial $^{87}Sr/^{86}Sr$ vs. wt% SiO_2. 135

Figure 20. Results of AFC modeling of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Calculations for AFC modeling based on expressions in DePaolo (1981). **(A)** - Nd (ppm) vs. $\varepsilon Nd_{(1895)}$ showing the results of varying proportions of assimilants (Archean tonalitic gneisses and metasedimentry Tasiuyak gneiss). **(B)** - Sr (ppm) vs. $^{87}Sr/^{86}Sr_{(1895)}$ showing selected values of "r" for different assimilants (Archean tonalitic gneiss and metasedimentary Tasiuyak gneiss). Ma/Mc = "r", where Ma = mass of the assimilant, and Mc = mass of fractional crystallization. "Kd" is the Nd and Sr bulk distribution coefficients (Arth, 1976; Reiners et al., 1996). 136

Figure 21. Cartoon cross-sections illustrating the proposed Early Proterozoic evolution in northern Labrador. (A): Basaltic underplating in the Nain Province margin in the period 2400-2200Ma (possibly accompanied by rifting) and associated dyke-emplacement and volcanism. Differentiation of the basaltic magma produces anorthosites. (B): During convergence between Nain and Rae provinces, subduction-generated magmas intrude the Nain margin in the period 1910-1864 Ma. Plutons emplaced through mafic underplated material record little or no contamination from continental crust (Group 1), whereas magmas emplaced outside or in the margin of the mafic underplated material record moderate to strong crustal contamination. The approximate location of the future Komaktorvik shear zone (KSZ) is indicated. (C): Anorthosites and amphibolites are emplaced as sheets within the KSZ by thrusting during Nain-Rae collision at ca. 1860 Ma. 140

Figure 22. Schematic cross-section showing variations in the volume of arc magmatism across the northern T.O.. By volume, the larger plutons interact comparatively less with host rocks (represented by gray zones) than smaller plutons. Thus Group 1 magmas are relatively uncontaminated by the Archean basement, whereas groups 2 and 3 are more strongly contaminated by Archean basement gneisses. 141

Figure 23. Schematic cross-section of the relation between crustal scale tilting and magma-host rock interaction in the northern Torngat Orogen. At the present erosion levels, the deepest levels are exposed in the eastern part of the tilted block. Assuming that magmas interact with their host rocks on their way up through the crust, magmas at lower crustal levels will record less contamination than those at higher crustal levels. In this way, the eastern plutons will be less contaminated, whereas the central and western ones will have experienced more interaction with crustal material. 143

CHAPTER 4

Figure 1. Precambrian tectonic elements of Laurentia from Hoffman, 1988. The Baltic shield is shown in a pre-Iapetus reconstruction., and Greenland is restored to rifting from North America. Uppercase names are Archean provinces; lowercase names are Proterozoic and Phanerozoic orogens. Abbreviations: BH, Black Hills inlier,; BL, Belcher belt; CH, Cheyenne belt; CS, Cape Smith belt, FR, Fox River belt; GL, Great Lakes tectonic zone; GS, Great Slave Lake shear zone; KL, Killarney magmatic zone; KP, Kapuskasing uplift; KR, Keweenawan rift zone; LW, Lapland-White Sea tectonic zone; MK, Makkovikorogen; MO, Mistassini-Otish basins; MRV, Minnesota River Valley terrane; SG, Sugluk terrane; TH, Thompson belt; TS, Transscandinavian (Smaland-Varmland) magmatic zone; VT, Vetrenny tectonic zone; WR, Winisk River fault. 152

Figure 2. Geological compilation map of Northeastern Laurentia from Van Kranendonk et al., 1993, showing inferred distribution of Archean cratonic blocks and Early Proterozoic orogens. 1= areas of unreworked Archean crust; 2= areas of reworked crust; 3= Early Proterozoic shelf-turbidite sequences; 4=Tasiuyak gneiss; 5= mafic volcanics; 6=Early Proterozoic arc rocks. Early Proterozoic supracrustal assemblages are as follows: A=Aillik Group; HB= Hoare Bay Group; I= Ingrid Group; K=Karrat Group; LH= Lake Harbor Group; M= Mugford Group; ML=Moran Lake Group; N= Narsajuaq arc; PC= Povungnituk and Chukotat groups; Pl= Piling Group; Pn= Penrhyn Group; R= Ramah Group; RG= Richmond Gulf Group; S= Sugluk Group; Sw= Seward subgroup; VS= Vallen and Sortis groups; W=Watts Group. ASZ= Abloviak shear zone; KSZ= Komaktorvik shear zone; WSZ= Wager shear zone. AIC= Ammassalik igneous complex. Pre-drift reconstruction from Rowley and Lottes (1988). 154

Figure 3. Hoffmans (1990) model for the tectonic assembly of Northeastern Laurentia. (A) Indentation of the Superior Province into Rae Province hinterland results in extrusion of an arm of the Rae Province to the southeast and formation of the ancestral Baffin basin. (B) Southerly thrusting of the Cape Smith beltin the Ungava orogenaccompanies collision between the southeast Rae and Superior provinces and emplacement of the Cumberland Batholith complex. (C) Accretion of Burwell and Nain Provinces. Modified according to Van Kranendonk et al., 1993, from Hoffman, 1990. 157

Figure 4. Schematic time-evolution diagrams fro the tectonic assembly of Northeast Laurentia between ca. 1.92-1.74 Ga from Van Kranendonk et al., 1993. Areas with horizontal lines are oceanic crust: small dots are platform sequences, turbiditic metasedimentary rocks and associated mafic rocks; large dots = Tasiuyak gneiss; x = sites of arc magmatism; + = late to post-tectonic granites; sutures = sites of collisional orogeny; white teeth = subduction zones; black teeth = thrusting; arrows indicate shear direction; circled C= Cumberland batholith complex. 159

Figure 5. Sm/Nd isotope data for Early Proterozoic metaigneous rocks from northern and southern segments of the Torngat Orogen, Northern Labrador. $\varepsilon_{Nd\ (1895)}$ vs. $^{147}Sm/^{144}Nd$ diagram for samples from Group 1, 2 and 3 from the northern TO (Chapter 3 of this study) and for eastern and western metaplutonic suites from the southern TO (Theriault and Ermanovics, 1997). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). Stippled field includes data from Nain and Rae Province Archean gneisses of the northern TO (Chapter 2). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). 162

Figure 6. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. Also shown are the location of Figure 2 ("This study") and the area described by Theriault & Ermanovics (1997). 164

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 Introduction

The Early Proterozoic is known to have been a time of intense orogenic activity and crustal accretion along Archean cratonic margins in North America, Greenland and the Baltic (e.g., Hoffman, 1988b; Bridgwater et al., 1991; Fig. 1). This is well illustrated in the Torngat Orogen, Northern Labrador, Canada, where a well-exposed series of Early Proterozoic and Archean rocks offer excellent opportunities to study crustal generation and collisional tectonics during the Early Precambrian (Figs. 2 and 3). The Torngat Orogen represents a mid-crustal section through a zone of collision between two Archean cratons, the Nain Province to the east, and the Rae Province to the west. Immediately prior to collision, the orogen was intruded by Early Proterozoic plutonic rocks.

The present study focuses on the geochemical and isotopic compositions of the Archean basement gneisses and the Early Proterozoic intrusive rocks. The aim is to document the petrogenesis of these metaigneous rocks and, in turn, to constrain the pre-collisional geometry of the orogen and the tectonic settings in which the rocks formed.

This chapter should be viewed as a framework and background for the three following chapters. Problems in Precambrian geology are introduced first, followed by outstanding questions concerning tectonic development of the TO, the goals of this thesis research and a summary of previous work in the Torngat Orogen. Following that, the field and laboratory methodology employed in this study is reviewed, followed by brief summaries of the three main chapters.

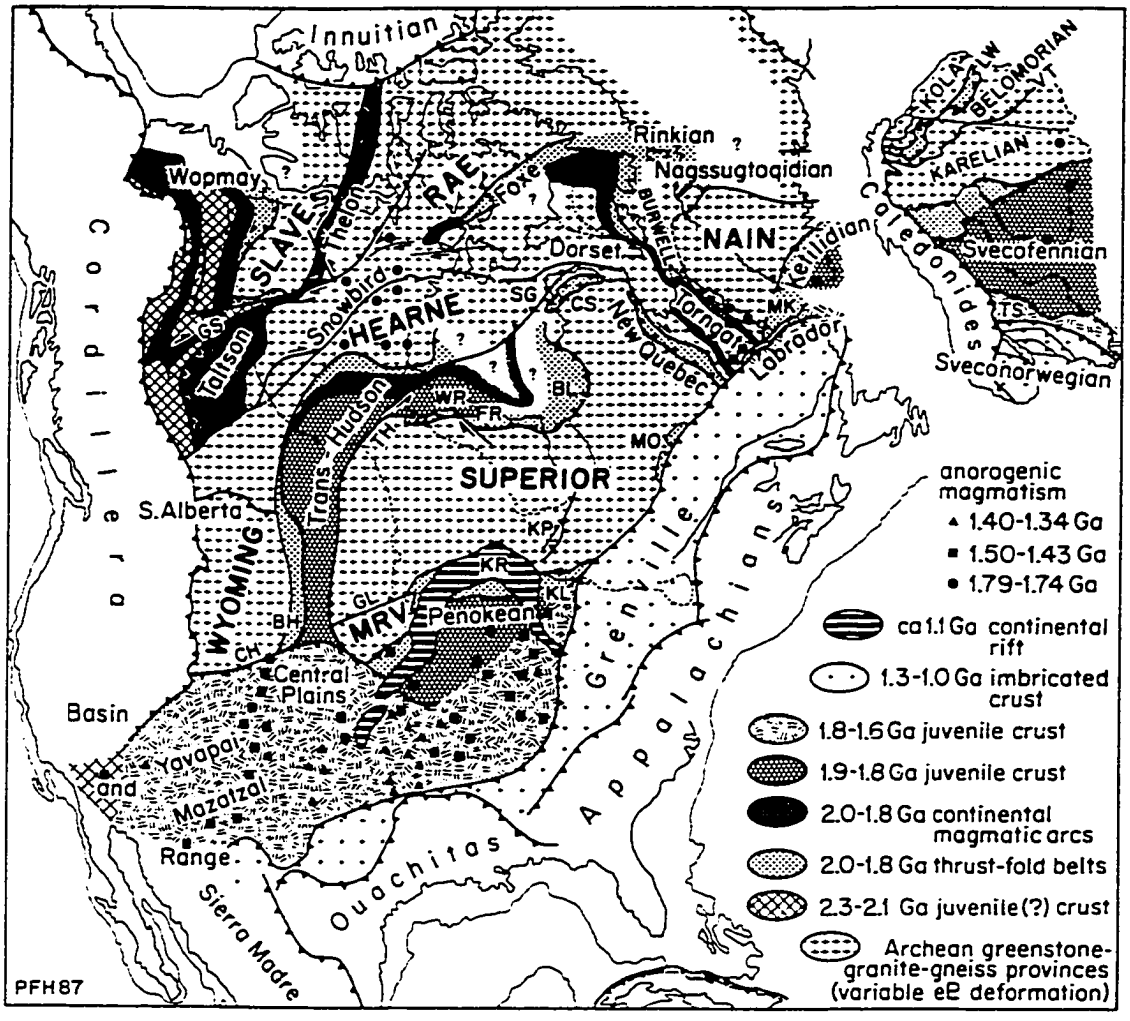


Figure 1. Precambrian tectonic elements of Laurentia from Hoffman, 1988. The Baltic shield is shown in a pre-Iapetus reconstruction.. and Greenland is restored to rifting from North America. Uppercase names are Archean provinces; lowercase names are Proterozoic and Phanerozoic orogens. Abbreviations: BH, Black Hills inlier,: BL, Belcher belt; CH, Cheyenne belt; CL, Cape Smith belt, FR, Fox River belt; GL, Great Lakes tectonic zone; GS, Great Slave Lake shear zone; KL, Killarney magmatic zone; KP, Kapuskasing uplift; KR, Keweenawan rift zone; LW, Lapland-White Sea tectonic zone; MK, Makkovikorogen; MO, Mistassini-Otish basins; MRV, Minnesota River Valley terrane; SG, Sugluk terrane; TH, Thompson belt; TS, Transscandinavian (Smaland-Varmland) magmatic zone; VT, Vetrenny tectonic zone; WR, Winisk River fault.

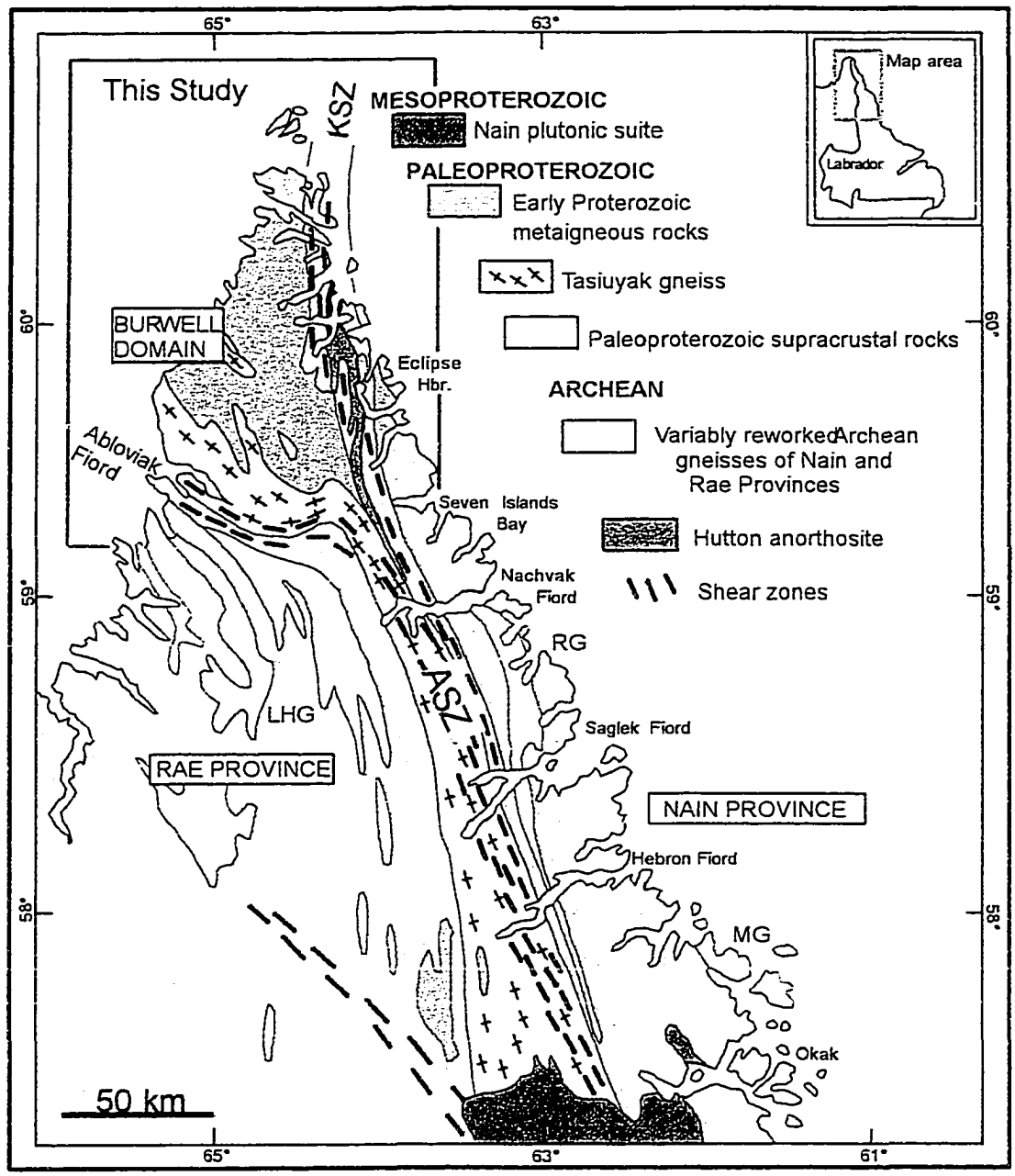


Figure 2. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. Also shown are the location of Figure 3 ("This study").

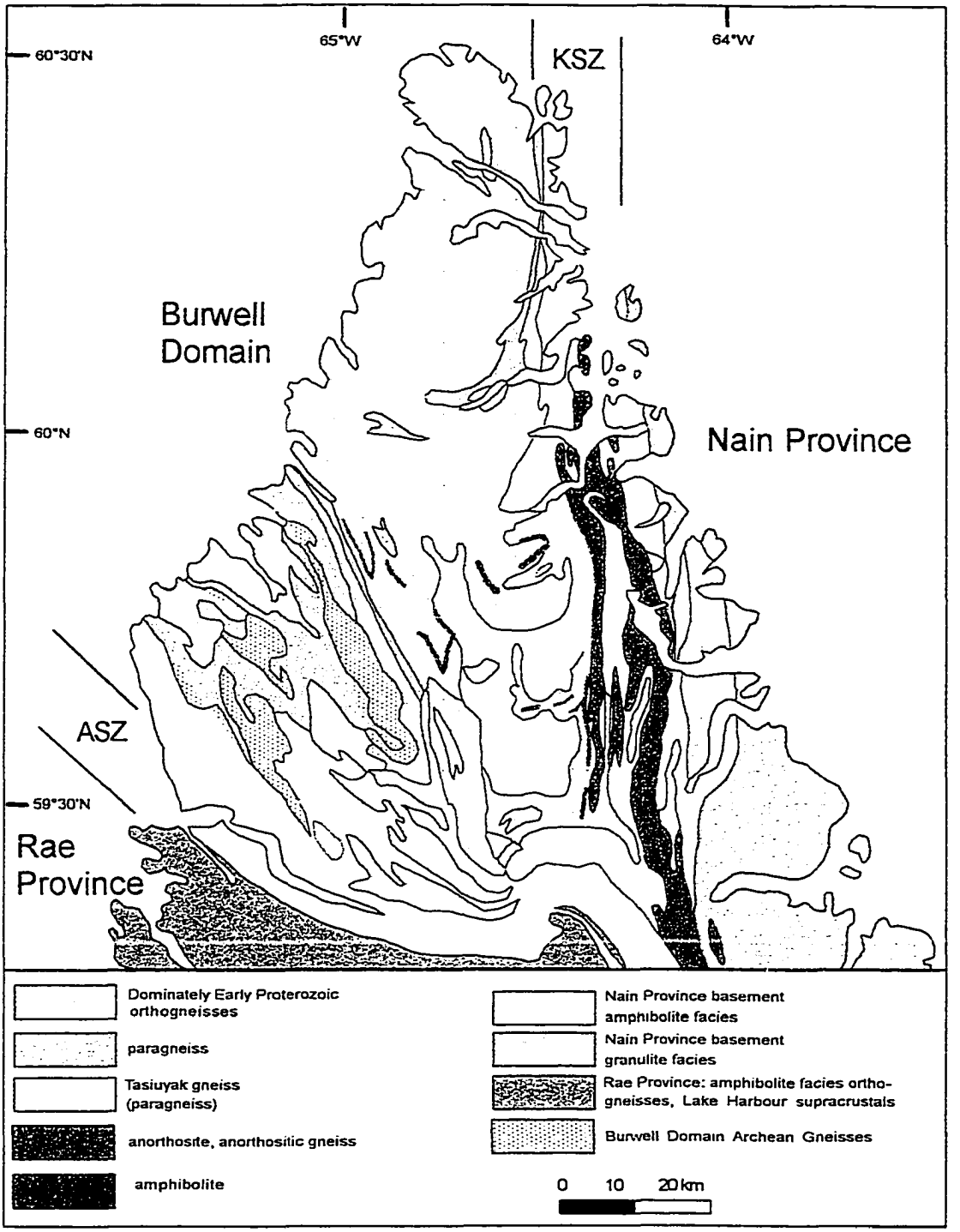


Figure 3. Geological sketch map of the study area in Northern Labrador (for location, see Fig.2) Based on data from Wardle et al. (1993), Van Kranendonk et al. (1994). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively.

1.2 Problems in Precambrian Geology

Generation of Archean crust appears to have been driven by significantly different processes than those driving crustal formation in the Proterozoic and Phanerozoic (e.g., Taylor and McLennan, 1985; Park, 1997). The evolution of Earth is driven by a complex interplay of processes which collectively work towards establishing internal equilibrium, primarily though the effects of cooling. At present, it is generally believed that by the end of the Archean, Earth had cooled sufficiently to sustain relatively rigid blocks of continental crust (i.e., "continents"), each of which represented an amalgam of fragments of older Archean crustal material (e.g., Park, 1997). These crustal blocks represent the onset of cratonization on Earth's surface, however the mechanisms which drove this Late Archean cratonization are still not fully understood. Most workers (e.g., Hoffman, 1988b; Park, 1997) now largely agree that by the beginning of the Proterozoic (ca. 2.5 Ga), Earth's crustal evolution had begun to operate largely by plate tectonic processes similar to those we observe today. This conclusion is based on: (1) the preservation of crustal features which suggest that the plates moved rigidly in response to contractional and extensional forces, and (2) the juvenile material added to pre-existing crust is lithologically similar to that forming in modern mid-ocean ridge, island arc and continental arc settings. Other tectonic features which are preserved in Early Proterozoic terranes and which have modern analogs include obducted ophiolites, rifted cratonic margins with associated sedimentary deposits, plutonic remnants of magmatic arcs and structural evidence for continent-scale collisional tectonics.

While it is now widely accepted that the Early Proterozoic marks the onset of Phanerozoic-style plate tectonics, there are still numerous questions concerning the factors which triggered this change on a global scale, and which processes drove Earth's evolution during the Archean (i.e., pre- 2.5 Ga). Such questions lay the foundations for most thematic studies in Precambrian geology, which include the temporal change in geochemical and isotopic characteristics of continental crust (e.g., Taylor and McLennan, 1985; Drummond and Defant, 1990; Martin, 1993); rates of crustal growth (e.g., McCulloch, 1993; McCulloch and Bennett,

1994; Bowring and Housch, 1995); crustal recycling (e.g., Bowring and Housh, 1995; McCulloch, 1993); modes of crustal addition (e.g., basalt underplating vs. arc magmatism vs. tectonic accretion of oceanic plateaus (e.g., Kay et al., 1992; Furlong and Fountain, 1991; Condie, 1997); and variations in subduction zone processes, such as slab vs. mantle wedge melting (Drummond and Defant, 1990; Martin, 1986, 1993; Wyllie, 1971).

Numerous studies have been directed towards constraining the evolution of crustal composition over time. For example, Taylor and McLennan (1985), used geochemical and isotopic analysis of sedimentary rocks of different ages to document that the average chemical compositon of crust formed in the Archean was different from that formed in the post-Archean. These changes include enrichments in large ion lithophile elements (LILEs), development of negative Eu anomalies and depletion in Ni, Co, Zr and Nb in post-Archean relative to Archean crustal rocks.

Much of the preserved Archean continental crust on Earth occurs in high-grade Archean terranes dominated by tonalites, trondhjemites and granodiorites (TTG; also referred to as tonalites, trondhjemites and dacites (TTD), where dacite is the extrusive equivalent of granodiorite). TTGs are characterized by a unique set of geochemical signatures, including strongly fractionated rare earth elements (REEs), high Sr/Y, Al_2O_3 > 15 wt.% and low K_2O/Na_2O (< 0.5; e.g., Barker, 1979; Drummond and Defant, 1990; Martin, 1993). The majority of petrogenetic studies of TTGs have concluded that the rocks formed by partial melting of juvenile (mantle-derived) basalt which was previously metamorphosed to garnet-amphibolite or eclogite (e.g., Martin, 1993; Drummond and Defant, 1990; Arth, 1979; Barker and Arth, 1976).

Martin (1986, 1993) and Drummond and Defant (1990) have, among others, suggested that the transition in geochemical character of average continental crust at the Archean-Proterozoic boundary reflects a change in subduction zone processes from partial melting of the subducted slab to partial melting of the overlying mantle wedge. Such a transition can be directly linked to changing thermal conditions of Earth. Earth's internal heat production during the Archean is

known to have been about three times the present values, due to the decay of short-lived radiogenic isotopes, and thus the total heat flux of Earth was greater by a comparable factor (Abbott and Hofmann, 1984). It is therefore not surprising that the tectonic and crust-forming processes during the Archean were different to those of today, and have evolved over time, in response to the near-exponential cooling of Earth.

The decrease in the heat-production rate has resulted in slower rates of overturn in the mantle throughout Earth history (Bickle, 1978). At present, over 65% of Earth's heat loss results from the formation of new oceanic crust at seafloor spreading centers (Sclater et al., 1981). Because this is the most efficient way to dissipate heat from Earth's interior, it is argued that similar processes (i.e., mantle convection) were also the most effective way to disperse heat during the Archean. This was likely achieved via more rapid generation and recycling of oceanic crust. In order to accomplish this, it has been calculated that in order to dissipate three times the heat of today, an 18-fold increase in ridge length would be required, and consequently the maximum age of subducted Archean lithosphere would be < 20 Ma, relative to ca. 60 Ma today (Hargraves, 1986; McCulloch, 1993).

On the basis of such considerations, Martin (1986; 1993), Drummond and Defant (1990), McCulloch (1993), and others, proposed that during the Archean, shallow subduction of buoyant, hot and young (< 20 Ma) oceanic crust resulted in partial melting which produced "high-Al" (Al_2O_3 > 15 wt.%) TTG crustal suites with garnet-amphibolite to eclogite residues. McCulloch (1993) argues that during the Archean, partial melting of the subducted slab took place in the upper 200 km of the mantle, and that subsequent dis-aggregation of the residue occurred in the upper 200–400 km. Rapid convection within this region of the upper mantle was isolated from the lower mantle by a high-viscosity transition zone (400-650 km) and as a result of this isolation, the bulk of Archean continental crust was probably extracted from the upper region.

Although the formation of crust by accretion of magmatic arcs is considered by many to be the primary mode of continental crustal growth today (e.g., Taylor and McLennan, 1985), it is

argued by some workers that this process alone could not account for the volume of crust produced by the end of the Archean (e.g., Kay et al., 1991). It has been proposed that underplating of mantle-derived basalt at the base of the crust followed by partial melting of this basalt to form tonalitic and trondhjemitic melts is another viable method of crustal addition (Furlong and Fountain, 1991; Bohlen and Mezger, 1989; Kay et al., 1992). Unlike accretion along magmatic arcs, this process can occur within-plates as well as at plate margins (Kay et al., 1992). If such a mode of crustal addition were active during the Archean, it could in part account for the large volumes of tonalitic and trondhjemitic crust observed, however in order to invoke this process to explain large volumes of Archean "high-Al TTD" suites, it is necessary to account for the fate of even larger volumes of the associated garnet-amphibolite or eclogite residues. Although widespread occurrences of post-Archean mafic xenoliths (representing the lower crust) have been documented (Rudnick, 1992), these are somewhat rare in Archean terranes and the evidence from exposed crustal cross-sections is that the lower part of the Archean continental crust does not contain a significant volume of garnet-amphibolite or eclogite (e.g., Percival et al., 1992).

A partial solution to this problem was proposed by Turcotte (1989) and Kay et al. (1992), who suggested that basalt is underplated at the base of the crust in regions where the crust has been doubly thickened by collision (i.e. > ca. 35 km). At this depth, basalt is transformed to garnet-bearing amphibolite. If partial melting and separation of a tonalite-trondhjemite fraction occurs, then the residue and its adjacent lithospheric mantle will be denser than the underlying, hot asthenosphere and will delaminate from the overlying crust. In this way, the tonalite-trondhjemite melt component of an originally basaltic crust is preserved, while the more mafic residue is recycled into the mantle.

In the two models discussed here, the processes required to generate the "high-Al TTD" suites preserved in Archean terranes include production of a mantle-derived tholeiitic basalt which was subjected to a second-stage partial melting. In both models, a residue composed of garnet-amphibolite or eclogite must be recycled back into the mantle, since large volumes of these

lithologies are not preserved in the Archean terranes, nor have they been documented to exist by seismic studies, by analysis of lower crustal xenoliths or in exposed lower crustal sections (Durheim and Mooney, 1994; Rudnick, 1992; Percival et al., 1992).

The formation of crust appears to have slowed since the Archean and it is estimated that by the beginning of the Proterozoic, somewhere between 50-95% of the present volume of continental crust had already been formed (Taylor and McLennan, 1985; McCulloch and Bennett, 1994). It is, however, unclear how much of this early crust has been recycled back into the mantle, and how much has been re-introduced into the crust as later juvenile additions through partial melting of the mid to lower crust.

Despite the remarkable technical and analytical advancements in the fields of geochronology, isotope geochemistry and geophysics, there are many unresolved issues such as those discussed above, concerning the mechanisms which drove crust-generation and amalgamation of continental masses during the Precambrian. Exposed Late Archean-Early Proterozoic terranes like the Torngat Orogen in Northern Labrador, Canada, offer unique opportunities to investigate this critical phase of Earth history in the field, and to study the transition in both tectonic and crust-formation processes which occurred at the Archean-Proterozoic boundary.

1.2 TORNGAT OROGEN - PRESENTATION OF STUDY AREA

1.2.1 Introduction

The Torngat Orogen (TO) preserves rocks which record segments of Earth history ranging between ca. 3.8 through 1.6 Ga, and include rocks formed in both extensional (2.3 - 2.2 Ga; Ermanovics, 1993; Connelly and Mengel, 1996) and convergent settings (ca. 1.91 - 1.86 Ga; e.g., Scott, 1995a). The TO further records continental collision and crustal thickening (ca. 1.86 Ga; Bertrand et al., 1993), resulting in fusion of two Archean cratons.

1.2.2 Outstanding questions related to tectonic development of the Torngat Orogen: status at the start of the project

In 1991, a three-year mapping project lead jointly by the Geological Survey of Canada (GSC) and the Newfoundland Department of Mines and Energy (NDME) was initiated in order to provide geological coverage of the northern segment of the Torngat Orogen (north of 59° 15', Figure 2). Prior to initiation of the project, it was already apparent that significant geological differences existed between the northern and southern segments, for example the divergence between the Komaktorvik and Abloviak shear zones and the anomalous field characteristics of the Burwell domain relative to adjacent Nain and Rae provinces (cf. previous work described below: Figs. 2 and 3).

At the onset of the project, the Torngat Orogen was recognized as the deeply eroded remnant of a zone of continental collision between the Archean Nain and Rae provinces (Figure 2). The Abloviak shear zone and associated metasedimentary Tasiuyak gneiss were considered to mark a crustal-scale suture, and deformation with localized mylonitization of Nain Province gneisses was documented along the Komaktorvik zone in the northern part of the Orogen (Figure 2).

Among the outstanding questions were the significance of the Komaktorvik zone as either a lithospheric zone of weakness within the Nain Province, or as a suture separating the western edge of the Nain Province from an independent craton, the Burwell domain. However, the Burwell domain was also poorly understood. It was not known if it represents an Archean or Early Proterozoic crustal block and its linkage with the adjacent Nain and Rae provinces was not constrained. The metamorphic history and the timing of discrete deformational events were also poorly constrained in this northern region. In addition to these factors, Early Proterozoic metaigneous rocks which occur throughout the area had not yet been identified. They were first documented during the initial (1991) field season (Van Kranendonk and Scott, 1992; Wardle et al., 1992), and thus their composition, modes of emplacement, relative volumes and spatial distribution across the northern Torngat region were major questions addressed during subsequent field seasons

(1992-1993). Once identified, the existence of these rocks precipitated many other significant questions concerning the tectonic evolution of the region. For example, did the Early Proterozoic igneous rocks form as an arc, and if so, what constraints does this allow us to place on the geometry of this collision zone?

In order to address these questions, integrated studies were carried out in the fields of structure, metamorphism, geochronology, igneous petrology and isotope geochemistry during the 1991-1993 field seasons. The results from these studies on the timing and orogenic development of the TO are summarized below in section 2.4.

1.2.3 Overview of previous geological studies in the Northern Torngat Orogen

Prior to initiation of this project, the most complete geological investigation of the northern Torngat region was conducted by the Geological Survey of Canada and led by Fred Taylor from 1967-1971 (Taylor, 1979). Taylor interpreted the area of northeastern Quebec and northern Labrador as consisting of two Archean blocks, the Nain Province to the east and the Superior Province to the west, separated by the Churchill Province, a broad region of Early Proterozoic rocks (subsequently called the Southeast Rae Province by Hoffman (1988b)). In Taylor's interpretation, the Nain and Superior Provinces preserve independent pre-2.6 Ga histories, but were both overprinted by high grade metamorphism at ca. 2.6 Ga (the Kenoran orogeny). Taylor documented an eastward increase in Proterozoic deformation across the Churchill Province, culminating in mylonitization of an extensive package of metasediments (later named Tasiuyak gneiss by Wardle (1983)), but did not report any Proterozoic magmatism in the northern Torngat region.

Subsequent to Taylor's regional mapping, geological investigations in the Torngat region were limited to the southern and central segments. Interest in the region was intensified as its significance as a continental boundary became clear. Mapping projects by the Newfoundland Department of Mines and Energy in the areas of Saglek, Hebron and Nachvak Fjords (Ryan et al.,

1983, 1984; Wardle, 1983, 1984; Mengel, 1984, 1985). Geological Survey of Canada in the Okak area (Ermanovics et al., 1989; Ermanovics and Van Kranendonk, 1990) and related structural and metamorphic studies (e.g., Mengel, 1988, Mengel & Rivers, 1990; Van Kranendonk and Ermanovics, 1990; Mengel et al., 1991), were aimed at determining the architecture and evolution of the Torngat Orogen. These studies formed the basis for establishing and defining the overall lithotectonic elements and main structures in the TO.

Mengel (1985, 1988) identified and interpreted the kinematic evolution of the Abloviak shear zone and developed the general model for tectonic collision that formed the TO. Mengel & Rivers (1991) constrained the P-T-t history of the orogen as preserved in the deformed western margin of the Nain Province and adjacent Tasiuyak gneisses in the Abloviak shear zone. The study documented crustal thickening to double normal thickness (10-11 kb, 800-850°C), followed by the development of a sinistral plastic shear zone at peak metamorphic conditions. Slip on the shear zone continued as it was exhumed 15-18 km, as documented by 5-6 kb of decompression and 150-200°C cooling which was followed by east-vergent crustal thickening along thrust faults. The displacement of deformed Nain Province gneisses (parautochtonous) and of allochthonous Tasiuyak gneisses (relative to the Nain Province) due to transcurrent shearing and unroofing led Mengel (1988) and Mengel et al. (1991) to conclude that these features developed in response to oblique collision between Nain and Rae cratons, and resulted in the welding of two Archean crustal blocks into a single continent.

Korstgaard et al. (1987) compiled the work of Ryan et al. (1983, 1984), Wardle (1983, 1984), Mengel (1984, 1985) and Korstgaard (1979a, b) into a coherent picture of the deformation and metamorphism associated with Early Proterozoic orogenesis in northern Labrador and West Greenland (Nagssugtoqidian Mobile Belt). Similar to Bridgwater et al. (1973) and Mengel (1985), Korstgaard et al. (1987) pointed out the many similarities in the structural and metamorphic evolution in the two regions, and further described how suites of mafic dykes intruded the Archean craton prior to Proterozoic orogenesis, and subsequently recorded the spatial distribution of

associated tectonic and thermal overprints recorded in the dykes. Korstgaard et al. (1987) named the north-trending zone of intense deformation within Nain Province gneisses along the western margin of the Nain Province the Komaktorvik zone, and noted its transition into mylonitized, granulite facies metasedimentary Tasiuyak gneisses in the Abloviak shear zone to the west (Fig. 2). The Abloviak shear zone had already been described as a zone of major transcurrent movement (Morgan, 1975; Bridgwater et al., 1973; Wardle, 1983, 1984), and its sinistral nature was defined by Mengel (1985). Later studies, including Korstgaard et al. (1987), described the tectonic significance of the Abloviak shear zone, and considered it a suture that records a major lithological break between two previously independent Archean cratons: the Nain and Churchill Provinces.

Korstgaard et al. (1987) proposed the name "Burwell domain" for the ambiguous, triangular shaped segment of crust bounded between the Komaktorvik and Abloviak zones in the northernmost region of the TO (Fig. 2). Though previously interpreted as a segment of Early Proterozoic Churchill province by Taylor (1979), Korstgaard et al. (1987) suggested that the Burwell domain represented a westward continuation of the Archean Nain Province, based mainly on its aeromagnetic signature.

Concurrent with the more detailed studies described above, Hoffman (1988a, b, 1989) considered the Torngat Orogen in terms of its significance during Proterozoic amalgamation of North America. Hoffman (1988a, b) subdivided the Churchill Province and named the segment west of the TO the SE Rae Province. In accord with evidence for east-vergent thrusting along and east of the Komaktorvik shear zone (Taylor, 1979; Wardle, 1983; Mengel, 1985; Korstgaard et al., 1987), Hoffman speculated that the SE Rae Province was thrust eastward over the western margin of the Nain Province during subduction-driven continental collision and suggested that if an ocean of significant width had been consumed by subduction, then related magmatism of Early Proterozoic age should occur in the Rae hinterland west of the Abloviak shear zone. In the North River map area (southern TO, Figure 2), Ermanovics et al. (1989), Ermanovics and Van Kranendonk (1990) and Van Kranendonk and Ermanovics (1990) identified the Lac Lomier

(during Torngat orogenesis) eastern margin of the Rae Province and found it to include Early Proterozoic metaplutonic "charnockites", thus confirming Hoffman's predictions. Compositionally similar charnockites were also identified in the adjacent Tasiuyak gneiss, and it was therefore argued that the Rae Province and Tasiuyak domain were juxtaposed prior to magmatic emplacement at ca. 1877 Ma (Bertrand et al., 1993). Ermanovics and Van Kranendonk (1990) and Van Kranendonk and Ermanovics (1990) further described the structural evolution of the area through detailed documentation of progressive Early Proterozoic deformation in a sinistral, transpressive regime, including a component of underthrusting of the Nain Province. The charnockites in the Tasiuyak gneiss and Lac Lomier complex were speculated to represent deep-level magmatic products of this subduction. Thus, based on the geology of the North River area, many workers adopted the model of westward-dipping subduction proposed by Hoffman (1988a, b, 1989) and applied it to the entire TO. This model, however, became difficult to reconcile with the geological features which were subsequently observed in the northern segment of the TO during the 1991-1993 mapping project (see below).

The following summary of results from the 1991-1993 mapping project is provided for the reader in order to clarify the general geology and timing of Early Proterozoic events in the northern TO, with which the present work is integrated. In the the northern TO an extensive suite of Early Proterozoic plutons intrude the Nain Province to the north and east of the ASZ. These Early Proterozoic rocks have been interpreted to represent a continental arc suite (e.g., Van Kranendonk & Scott, 1992; Scott & Campbell, 1993; Rivers et al., 1996; Van Kranendonk & Wardle, 1996).

Early Proterozoic calc-alkaline metaplutonic rocks as well as abundant supracrustal rocks occur within the Burwell domain, but the extent to which the area is underlain by Archean basement gneisses is not well known. The Komaktorvik shear zone (KSZ) is developed in deformed Archean gneisses, a belt of anorthosites (the Hutton anorthosite, which is commonly

juxtaposed with layered amphibolites), deformed Early Proterozoic mafic dykes, metasedimentary gneisses and Early Proterozoic magmatic rocks.

Integrated field and geochronological studies have established that in the northern TO, arc-magmatism occurred at 1910-1864 Ma (Scott, 1995a, Scott & Machado, 1995), followed by continental collision between the Archean Nain and Rae Provinces (D_1; 1870-1860 Ma, Bertrand et al., 1993; Scott & Machado, 1995; Scott, 1995a,b). During continued (oblique) convergence between Nain and Rae Provinces, the TO was cut by a network of crustal scale, brittle to ductile shear zones, of which the largest are the ASZ and KSZ. The ASZ is a sinistral, transcurrent shear zone, which developed during the period ca. 1845-1822 Ma (D_2; Bertrand et al., 1993). Movement along the KSZ was oblique-sinistral and was mainly active during ca. 1798-1780 Ma (D_3; Bertrand et al., 1993; Scott 1995a).

1.2.4 Thesis goals

This Ph. D. thesis presents the results of a geochemical and Nd, Sr and Pb isotope study of Archean and Early Proterozoic rocks from the Torngat Orogen in northern Labrador, Canada (Figs. 1 and 2). The overall aim is to characterize the rocks in terms of their geochemical and isotopic composition and to use this information to better constrain (1) crust-forming processes by which they were generated; and (2) tectonic development of the northern TO.

The thesis is based on field studies which include lithological and structural mapping and sample collection carried out by the author as a participant in the regional mapping project in northern Labrador described above.

In detail, the main goals of this thesis can be broken down into several related objectives:

1) to establish the geochemical and isotopic composition of the main lithotectonic units that comprise the northern Torngat region,
2) to evaluate the isotopic and geochemical signatures of Archean basement gneisses on either side of the collision belt which comprises the TO, as a means to investigate (a) their

petrogenesis and (b) the linkages between the Archean Nain and Rae provinces and the Archean gneisses of the Burwell domain, to provide constraints on the distribution, geometry and pre-orogenic histories of individual Archean terranes,

3) to provide detailed geochemical and isotopic data on Early Proterozoic magmatic rocks across the region in order to (a) constrain the petrogenesis of these rocks, (b) establish the spatial distribution of key compositional characteristics, and (c) to incorporate the data from (a) and (b) into models for the tectonic development of the TO and relate these to Phanerozoic models, and

4) to integrate the findings above (points (1)-(3)) into larger-scale models for the tectonic evolution and assembly of NE Laurentia.

1.2.5 Field component of the project: Logistics

The field component of this Ph. D.-project was conducted in cooperation with a three-year mapping program run jointly by the Geological Survey Branch of the Newfoundland Department of Mines and Energy and the Geological Survey of Canada, with some support from the Eastern Canadian Seismic Onshore and Offshore Transect (ECSOOT) initiative under the Canadian LITHOPROBE project. The aim of the field program was to map the Archean and Early Proterozoic rocks of the Torngat Orogen in northernmost Labrador (Figure 2) and the primary objectives were to:

1) provide regional geological coverage at the scale 1:50000,

2) assess the economic potential of the area (i.e., mineralizations),

3) enhance the understanding of the geological evolution of the orogen through a multidisciplinary study integrating field-mapping, structural geology, metamorphic petrology, geochronology, isotope geochemistry and igneous petrology.

My role in this project was, as senior mapping geologist, to complete regional mapping and documentation of field data, to collect samples for geochemical and isotopic analyses during

the course of daily mapping traverses, and, with this as a basis, to carry out a geochemical and isotopic study of the main lithotectonic units that comprise the northern Torngat region.

1.3 METHODOLOGY

1.3.1 Introduction

One of the primary objectives of this thesis is to determine the origin of Early Proterozoic magmatic rocks occurring within the TO. This has significant ramifications for the tectonic development of the region which lead to formation of new crust during this phase of Earth history. A major question is whether Early Proterozoic magmatism was the result of mantle melting via subduction zone processes, or whether it represents partial anatexis of the Archean basement gneisses and supracrustal rocks, induced by basaltic underplating or by high grade metamorphism and deformation associated with continental collision.

Another objective is to identify the isotopic or geochemical character of the Nain and Rae Province basement gneisses, which is used to link the Archean crust of unknown affinity in the Burwell domain to either the Nain or Rae Provinces. This is necessary in order to determine the pre-orogenic distribution of cratons and subsequent geometry of orogenic elements along the plate boundaries.

In order to address these problems, we have employed the Nd, Sr and Pb isotope systems to characterize the rocks in the study area. Combined with major and trace element data, these isotope systems are particularly useful in petrogenetic studies, and in the following we will briefly discuss the use and applicability of each system.

1.3.2 Nd isotope system

The Nd isotope system has been used here (1) to identify the source(s) from which the Early Proterozoic magmatic rocks were derived, and (2) as a method of "fingerprinting" or identifying linkages between crustal units on the basis of isotopic characteristics, in this case by

attempting to correlate Archean blocks of unknown affinity (Archean rocks in the Burwell domain) with Archean rocks in better constrained settings (i.e., the Nain and Rae provinces) on the basis of isotopic similarities.

Systematic variations in the isotopic composition of Nd exist within the Earth due to: (1) radioactive decay of ^{147}Sm to ^{143}Nd; and (2) chemical fractionation of Sm from Nd during differentiation processes, for example partial melting of the upper mantle to produce continental crust (DePaolo, 1988; Farmer and Ball, 1993). The elements Sm and Nd are both light rare earth elements (REE) and are geochemically almost identical with the exception that Nd has a slightly larger ionic radius than Sm and thus is preferentially partitioned into the melt during partial melting of the upper mantle. This causes variations in the Sm/Nd ratios of evolving rocks, such that mafic residue will have higher Sm/Nd than the resulting melt. The range in of $^{147}Sm/^{144}Nd$ is very small among most terrestrial rocks (0.1-0.3: DePaolo, 1988), and this, combined with a very long half-life (106 Ga), results in a very narrow range of $^{143}Nd/^{144}Nd$ values on Earth. It has therefore become conventional to express these differences in terms of "εNd", where:

$$\varepsilon Nd\ (T) = (((^{143}Nd/^{144}Nd_{(rock)} / {}^{143}Nd/^{144}Nd_{(CHUR)}) - 1) * 10^4$$

where $^{143}Nd/^{144}Nd_{(CHUR)}$ refers to the $^{143}Nd/^{144}Nd$ value in a model chondritic uniform reservoir (CHUR), which is based on the measured values in chondritic meteorites (DePaolo, 1988).

Based on this convention, Archean rocks with felsic to intermediate compositions have εNd values which are strongly negative, reflecting the long term light rare earth element (LREE) enrichment of continental crust relative to the bulk Earth or to CHUR (Farmer and Ball, 1993). In contrast, mafic rocks derived from depleted mantle (i.e., mantle from which crust forming elements have been extracted over time), have positive εNd values (typically between 0 and +10; Farmer and Ball, 1993), making discrimination between these two end-member sources very distinct. The strongly divergent isotopic compositions between Archean felsic continental crust vs. mafic crust

(e.g., oceanic crust) is attributed to the fact that the relatively LREE enriched Archean crust has been isolated from its LREE-depleted upper mantle source for a long time. The Nd isotopic evolution of the depleted upper mantle (DM) has been approximated by determining initial isotopic compositions of mafic crustal rocks of various ages which were derived from this source (DePaolo, 1988).

At the time that igneous rocks separate from their source, they will have the same $^{143}Nd/^{144}Nd$ (but a different $^{147}Sm/^{144}Nd$). Therefore, if the crystallization age, the measured $^{143}Nd/^{144}Nd$ value, and Nd and Sm concentrations are known for a rock, it is possible to calculate the $^{143}Nd/^{144}Nd$ value that the rock had at the time it was formed, and thus to identify the $^{143}Nd/^{144}Nd$ value of its source. On the basis of this information, it is then possible to identify whether the rock was derived from deformed and partially melted ('reworked') older continental crust or from depleted mantle.

Using this technique (and assuming that the system has remained closed to mobilization of Sm or Nd since crystallization of the rock), we can determine the initial isotopic compositions for a suite of rocks and identify spatial variations in these values which may indicate a change in the source(s) (i.e., derivation from a depleted mantle source, from Archean basement gneisses, or from a combination of these).

The Nd isotopic system is particularly useful in this study for two reasons: (1) The rocks comprising the TO have experienced relatively high grade metamorphism, which is commonly accompanied by mobilization of certain elements (e.g., Touret, 1996). Because Nd and Sm (the parent and daughter elements in this isotope system) have similar chemical characteristics (+3 valence and approximately similar ionic radius), Sm and Nd behave similarly during most secondary processes, and are less likely to be fractionated from each other during such processes unless partial melting is involved (DePaolo, 1988; Bridgwater et al., 1989). (2) At ca. 1895 Ma (the approximate igneous age of the Early Proterozoic rocks; e.g., Scott, 1995a), the Nd isotopic composition of these two likely end-member sources (i.e., Archean crust comprising the northern

TO and underlying depleted mantle), are very distinct (approximately -6 to -18 and +5, respectively).

1.3.3 Sr isotope system

In contrast to the Sm-Nd isotope system, the parent-daughter elements Rb and Sr are significantly different from each other in their chemical characteristics. Both Rb and Sr are lithophile, incompatible elements, but Rb is an alkali element with a +1 valence whereas Sr is an alkaline earth element with a valence of +2. Thus the two elements have substantially different mineral affinities in both mantle and crustal fractionation processes (Faure, 1986). As a result, there are large variations in terrestrial Rb/Sr whole rock ratios. This would potentially lend the Rb/Sr system to providing very distinctive fingerprints for particular source characteristics and igneous processes and would provide a good spread for Rb-Sr isochron ages. However, because of the highly mobile behavior of Rb in response to metasomatism and metamorphism, and because radiogenic ^{87}Sr readily diffuses during even low grade metamorphism, the Rb-Sr system is generally subject to open system behavior (e.g., Faure, 1989; Bridgwater et al., 1989). This is particularly true in high grade metamorphic rocks such as those in the TO, and thus Rb-Sr data must be interpreted carefully.

Disruption and homogenization of an isotope system can, however, also provide useful information, particularly when used in combination with other isotope systems. For example, when rocks experience high grade metamorphism, the Nd isotopic system will commonly remain closed to diffusion, whereas the Rb-Sr system may be entirely homogenized and reset. If the Rb-Sr system subsequently remains closed, it will record the age and nature of the metamorphic event. Such open system behavior can also record information about the type of secondary processes, for example Rb loss during granulite facies metamorphism, or Rb gain during retrograde metamorphism (e.g., Bridgwater et al., 1989; Touret, 1996).

In this study, we use the Rb-Sr isotope data in conjunction with Nd isotope data and U-Pb geochronological data. In Chapter 2 it is used primarily to identify the time integrated effects of Rb-loss associated with Archean granulite facies metamorphism and to document how this differs between the Archean provinces. In Chapter 3 it is used as a general guide for source compositions of the Early Proterozoic metaigneous rocks, and for identifying how these vary spatially across the TO.

1.3.4 Pb isotope system

The $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ isotopic method keeps a time-integrated record of the U/Pb values of rocks or minerals and is insensitive to recent U loss. The system differs from other isotopic methods in that it combines two independent decay systems (^{238}U to ^{206}Pb and ^{235}U to ^{207}Pb), thus eliminating the need to measure U (Faure, 1986; Gariépy and Dupré, 1991). However, similar to the Rb-Sr isotope system, the Pb-Pb isotope system is limited in terms of its use as a direct geochronometer, or as a means to constrain initial isotopic compositions. This is due to the extreme mobility of both parent and daughter elements during most secondary processes. Nonetheless, Pb isotopes provide an extremely useful tool for tracing crust-mantle evolution through time.

For example, during partial melting of the mantle, U and Th are highly incompatible, with Th being somewhat more incompatible than U. Pb is also incompatible, but to a lesser extent than U or Th, and thus the U/Pb and Th/Pb values of the resulting magmas are higher than those of their mantle source (Faure, 1986; Gariépy and Dupré, 1991).

The large variations observed in Pb isotope compositions on Earth reflect the differences in geochemical behavior of U, Th and Pb during terrestrial differentiation processes integrated over time (DePaolo, 1988; Gariépy and Dupré, 1991). This has resulted in the development of isotopic reservoirs, or regions of Earth that share similar isotopic compositions due to the fact that they have undergone similar geological differentiation histories (Gariépy and Dupré, 1991), however,

because of commonly occurring mixing between these reservoirs. the isotopic compositions of the reservoirs can also be heterogeneous.

Pb isotopes serve as a particularly sensitive method for detecting crustal contamination in intrusive rocks. Because Pb concentrations in the crust are far greater than those in (mantle derived) melts. the isotopic compositions of the melts are displaced towards values of the crust through which they rise. to a larger degree than what is observed in Sr or Nd isotope compositions (Bridgwater et al.. 1989: Gariépy and Dupré. 1991).

High grade metamorphic terranes (i.e.. lower continental crust). are commonly characterized by strong depletion of U and Th (e.g.. Moorbath et al.. 1969: Bridgwater. et al.. 1989: Gariépy and Dupré. 1991). There are numerous theories that account for this observation. including the removal of U and Th by fluid phases during granulite facies metamorphism or through magmatic events in the lower crust (e.g.. Touret. 1996).

The time-integrated outcome of such U and Th depletion is that old (Archean) granulite facies terranes commonly are characterized by unradiogenic Pb. suggesting that they have remained stable for sustained periods of time. Subsequent orogenies may rejuvenate the lower crust through additions of radiogenic Pb from the upper crust and mantle-derived magmas (Rudnick and Goldstein. 1990).

4. STRUCTURE OF THE THESIS

4.1 Introduction

This thesis is divided into four chapters of which Chapters 2. 3. and 4 will be modified into publishable papers that each discuss a particular aspect of the geological evolution of the Torngat Orogen and the effect this work has on constraining the Proterozoic assembly of northeast Laurentia. Each of these chapters is a stand-alone manuscript. so a certain amount of repetition is unavoidable. e.g.. regional setting. geology of the TO. references. etc.

The present introductory chapter contains the general geologic background for the study and outlines the scope and goals of the thesis. The contents of the other chapters are summarized below.

1.4.2 Chapter 2 - *("Isotopic and geochemical constraints on Archean crustal development prior to and during the Torngat Orogeny, N. Labrador")*

This chapter describes the geologic history of the Archean basement gneisses which were affected by Proterozoic convergence, arc magmatism and continental collision. It presents an isotopic and geochemical characterization of the Archean provinces which constitute northernmost Labrador, identifies how these characteristics vary between provinces, and, on the basis of these data, discusses possible linkages between the enigmatic Burwell domain and adjacent Nain and Rae provinces. Chapter 2 also contains a brief discussion of the processes of crust-formation and amalgamation which were active during the Archean. This part of the dissertation provides a foundation for subsequent chapters and also contributes the first comprehensive data-set from this region.

1.4.3 Chapter 3 - *("Petrogenesis of Early Proterozoic magmatic rocks in the Torngat Orogen, Northern Labrador, Canada: Isotopic and Geochemical constraints on Early Proterozoic arc magmatism and tectonic assembly")*

In this chapter, a detailed isotopic and geochemical characterization of the Early Proterozoic magmatic rocks that occur in the northern Torngat region is presented. Chapter 3 provides a general petrogenetic model for the origin of these rocks, which includes an evaluation of the spatial variations in initial Nd isotope compositions across the orogen (which do not agree with models established for Phanerozoic orogens). This chapter also includes an evaluation of the geochemical effects of granulite facies metamorphism and integrates the data from this study with

other observations (documented in the literature), in order to further constrain the overall Early Proterozoic tectonic development of the region.

1.4.4 Chapter 4 - *("Models for the growth of Laurentia: revisions based on investigations in the Torngat Orogen, northern Labrador, Canada")*

The last chapter summarizes recent models for Early Proterozoic amalgamation of the Torngat Orogen and NE Laurentia, and suggests modifications to these models on the basis of the findings from chapters 2 and 3 above. Chapter 4 further addresses the current ideas concerning tectonic processes which were operative during the Early Proterozoic. The underlying purpose of this work, which will be further developed by the author in the future (proposed post-doctoral research), is to tie together, on a global scale, the processes of crustal generation and amalgamation which operated during Early Proterozoic time. The Early Proterozoic is of particular interest because it marks a transition in Earth history from the Archean processes of crustal formation and amalgamation, to the better constrained processes that are operative today. This time period is also significant because it marks the beginning of intense orogenic activity across North America, the North Atlantic and the Baltic.

REFERENCES

Abbott, D.H., Hoffman, S.E., 1984. Archean plate tectonics revisited 1. Heat flow, spreading rate, and the age of subducting oceanic lithosphere and their effects on the origin and evolution of continents. Tectonics, 3/4: 429-448.

Barker, F., Arth, J.G., 1976. Generation of trondhjemitic-tonalitic liquids and Archean bimodal trondhjemitic basalt suites. Geology, 4: 596-600.

Bertrand, J.-M., Roddick, J.C., Van Kranendonk, M.J., Ermanovics, I. 1993. U-Pb geochronology of deformation and metamorphism across a central transect of the Early Proterozoic Torngat Orogen, North River map area, Labrador. Can. J. Earth Sci., 30: 1470-1489.

Bickle, M.J., 1978. Heat loss from the earth: a constraint on Archean tectonics from the relation between geothermal gradients and the rate of plate production. Earth Planet. Sci. Lett., 40: 301-315.

Bohlen, S.R., Mezger, K., 1989. Origin of granulite terrains and the formation of the lowermost continental crust. Science, 244: 326-329.

Bowring, S.A., Housh, T., 1995. The Earth's Early Evolution. Science, 269: 1535-1540.

Bridgwater, D., Escher, A., Watterson, J., 1973. Tectonic displacements and thermal activity in two contrasting Proterozoic mobile belts from Greenland. Phil. Trans. R. Soc. Lond. A273: 513-533.

Bridgwater, D. Marker, M., Mengel, F., 1991. The eastern extension of the Early Proterozoic Torngat orogenic zone across the Atlantic. *In* ESCOOT Lithoprobe Report 27: 76-91.

Bridgwater, D. Rosing, M., Schiotte, L., Austrheim, H., 1989. The effect of fluid controlled element mobility during metamorphism on while rock isotope systems, some theoretical aspects and possible examples. *In* Fluid movements-element transport and the composition of the deep crust. *Ed* D. Bridgwater, NATO ASI Series C, 281:277-298.

Condie, K.C., 1997. Contrasting Sources for Upper and Lower Continental Crust: The Greenstone Connection. Jour. Geol., 105:729-736.

Connelly, J. N., Mengel, F. C., 1994. Preliminary U-Pb data from Archean gneisses in a transect across the Komaktorvik Shear Zone, Northern Torngat Orogen. *In* ESCOOT Lithoprobe Report 45: 72-82.

Connelly, J. N., Mengel, F. C., 1996 The Archean backdrop to Paleoproterozoic tectonism in the Nagssugtoqidian and Torngat Orogens: new constraints from U-Pb geochronology. *In* ECSOOT Lithoprobe Report 57: 15-24.

DePaolo, D.J., 1988. Neodymium Isotope Geochemistry, Springer-Verlag, Pubs., 187pp.

Drummond, M.S., Defant M.J., 1990. A model for trondhjemite-tonalite-dacite genesis and crustal growth via slab melting: Archean to modern comparisions. Jour. Geophys. Res., 95/B13: 503-521.

Durrheim, R.J., Mooney, W.D., 1994. Evolution of the Precambrian Lithosphere: Seismological and geochemical constraints. Jour. Geophys. Res., 99/B8:15359-15374.

Ermanovics, I., 1993 Geology of the Hopedale Block, southern Nain Province, and the adjacent Proterozoic terranes, Labrador, Newfoundland. Geological Survey of Canada Memoir, 431; 161pp.

Ermanovics, I., Van Kranendonk, M.J., 1990. The Torngat Orogen in the North River-Nutak transect area of Nain and Churchill provinces. Geosci. Can., 17: 293-296.

Ermanovics, I., Van Kranendonk, M.J., Corriveau, L., Mengel., F., Bridgwater, D., Sherlock, R., 1989. The boundary zone of the Nain-Churchill provinces in the North River-Nutakmap areas, Labrador. Geological Survey of Canada, Paper 89-1C: 385-394.

Farmer G.L., Ball, T.T., 1993. Origin and significance of Nd isotopic provinces within the Precambrian crust of the conterminous United States. *In* The Geology of North America Vol. C-2, Precambrian : Conterminous U.S.. The Geological Society of America: 614-636.

Faure, G., 1986. Principles of Isotope Geochemistry. John Wiley and Sons, Pubs., 589pp.

Furlong, K.P., Fountain, D.M., 1991. Continental Crustal Underplating: Thermal Considerations and Seismic-Petrologic Consequences. Jour. Geophys. Res., 91/B8: 8285-8294.

Gariepy, C., Dupre, B., 1991. Pb isotopes and crust-mantle evolution. *In* Short course handbook on applications of radiogenic isotope systems to problems in geology. *Eds.* L. Heaman and J. N. Ludden. Mineral. Assoc. Canada, volume 19: 191-224.

Hargraves, R.B., 1986. Faster spreading or greater ridge length in the Archean? Geology 14: 750-752.

Hoffman, P.F., 1988a. The extent of Trans-Hudson orogen and subdivision of the Churchill province. *In* The Early Proterozoic Trans-Hudson Orogen of North America, *Eds.* J.F. Lewry and M.R. Stauffer: Geol. Survey of Canada Spec. Pap. 37: 15-40.

Hoffman, P. F., 1988b. United plates of America, the birth of a craton: Early Proterozoic assembly and growth of Laurentia. Ann. Rev. Earth Planet. Sci. 16: 543-603.

Hoffman, P. F., 1989. Precambrian geology and tectonic history of North America. *In* The Geology of North America - An overview. *Eds.* A. W. Bally and A. R. Palmer. Geological Society of America, The Geology of North America. A: 447-512.

Kay, R.W., Kay, S.M., Arculus, R.J., 1992. Magma genesis and crustal processing *In* Continental lower crust. *Eds.* D.M. Fountain, R.J. Arculus, R.W. Kay: 423-443.

Kay, R.W., Kay, S.M., 1991. Creation and destruction of lower continental crust. Geologische Rundschau, 80/2: 259-278.

Korstgaard, J. A., 1979a. Metamorphism of the Kangamiut dykes and the metamorphic and structural evolution of the southern Nagssugtoqidian boundary in the Itivdleq-Ikertoq region, West Greenland. Rapp. Gronlands geol. Unders. 89: 63-75.

Korstgaard, J. (editor), 1979b. Nagssugtoqidian geology. Rapp. Gronlands geol. Unders. 89: 146pp.

Korstgaard, J., Ryan, B., Wardle, R., 1987. The boundary between Proterozoic and Archaean crustal blocks in central West Greenland and northern Labrador. *In* Evolution of the Lewisian and Comparable Precambrian High Grade Terrains, *Eds.* R. G. Park and J. Tarney. Geol. Soc. Lond. Spec. Publ. 27: 247-259.

Martin, H., 1986. Effect of steeper Archean geothermal gradient on geochemistry of subduction-zone magmas. Geology, 14: 753-756.

Martin, H., 1993. The mechanisms of petrogenesis of the Archean continental crust-Comparison with modern processes. Lithos, 30: 373-388.

McCulloch, M.T., 1993. The role of subducted slabs in an evolving earth. Earth Planet. Sci. Let., 115: 89-100

McCulloch, M.T., Bennett, V.C., 1994. Progressive growth of the Earths continental crust and depleted mantle: Geochemical constraints. Geochim. Cosmochim. Acta, 58/21: 4717-4738.

Mengel, F. C., 1984. Preliminary results of mapping in the Ramah Group and adjacent gneisses south of Saglek Fiord, northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 84-1: 21-29.

Mengel, F. C., 1985. Nain-Churchill Province boundary: a preliminary report on a cross-section through the Hudsonian front in the Saglek Fiord area, northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 85-1: 33-42.

Mengel, F., 1988. Thermotectonic evolution of the Proterozoic-Archaean boundary in the Saglek area, northern Labrador. Ph. D. thesis, Memorial University of Newfoundland, 335 p.

Mengel, F., Rivers, T., 1990. The synmetamorphic P-T-t path of granulite facies gneisses from Torngat orogen, and its bearing on their tectonic history. Geosci. Canada 17: 288-293.

Mengel, F., Rivers, T., 1991. Decompression reactions and P-T-t paths from individual samples in high grade rocks, northern Labrador: implications for Early Proterozoic tectonic evolution. J. Petrol. 32: 139-167.

Mengel, F., Rivers, T., (in press). Metamorphism in the Paleoproterozoic Torngat Orogen, Labrador: Petrology and P-T-t paths of amphibolite- to granulite-facies rocks across the Komaktorvik shear zone. Canad. Mineral., 35: (1997).

Mengel, F., Rivers, T., Reynolds, P., 1991. Lithotectonic elements and tectonic evolution of Torngat orogen, Saglek Fiord, northern Labrador. Can. J. Earth Sci. 28: 1407-1423.

Moorbath, S., Welke, H., Gale, N.H., 1969. The significance of lead isotope studies in ancient, high-grade metamorphic basement complexes, as exemplified by the Lewisian rocks of northwest Scotland. Earth Planet. Sci. Lett., 6: 245-256.

Morgan, W. C., 1975. Geology of the Precambrian Ramah group and base-ment rocks in the Nachvak fiord-Saglek fiord area, North Labrador. Geol. Surv. Canada Paper 74-54: 42pp.

Park, R.G., 1997. Early Precambrian plate tectonics. S. Afr. J. Geol., 100/1: 23-35.

Percival, J.A., 1992. Archean tectonic setting of granulite terranes of the Superior Province, Canada: a view from the bottom. *In* Granulites and crustal evolution *Eds.* D. Vielzeuf, P. Vidal: 171-193.

Rivers, T., Mengel, F., Scott, D. J., Campbell, L. M., Goulet, N., 1996. Torngat Orogen - a Paleoproterozoic example of a doubly vergent compressional orogen. *In* Precambrian Crustal Evolution in the North Atlantic Regions. *Ed.* T. S. Brewer. Geol. Soc. Lond. Spec. Publ. 112, 117-136.

Rudnick, R.L., 1992. Xenoliths-Samples of the lower continental crust. *In* Granulites and crustal evolution *Eds.* D. Vielzeuf, P. Vidal: 269-313.

Rudnick, R.L., Goldstein, S.L., 1990. The Pb isotopic compositions of lower crustal xenoliths and the evolution of lower crustal Pb. Earth Planet. Sci. Lett., 98: 192-207.

Ryan, A. B., Martineau, Y., Korstgaard, J., Lee, D., 1984. The Archaean-Proterozoic boundary in northern Labrador, Report 2. *In* Current Research, part A, Geol. Surv. Canada Paper 84-1A: 545-551.

Ryan, B., Martineau, Y., Bridgwater, D., Schiotte, L., Lewry, J., 1983. The Archaean-Proterozoic boundary in the Saglek fiord area, Labrador: Report 1. *In* Current Research, part A, Geol. Surv. Canada Paper 83-1A: 297-304.

Sclater, J.G., Parsons, B., Jaupart, C., 1986. Oceans and continents: Similarities and differences in the mechanisms of heart loss, J. Geophys. Res., 86/11: 535-552.

Scott, D.J., 1995a. U-Pb geochronology of the Nain craton on the eastern margin of the Torngat Orogen, Labrador. Can. J. Earth Sci., 32/11: 1859-1869.

Scott, D.J., 1995b. U-Pb geochronology of a Paleoproterozoic continental magmatic arc on the western margin of the Archean Nain craton, northern Labrador, Canada. Can. J. Earth Sci., 32/11: 1870-1882.

Scott, D. J., Campbell, L. M., 1993. Evolution of the Paleoproterozoic Torngat Orogen, Labrador, Canada: Recent advances using U-Pb geochronology and Nd isotope systematics. *In* Geol. Soc. Am. Program with Abstracts, vol. 25, A-237.

Scott, D. J., Machado, N., 1995. U-Pb geochronology of the northern Torngat Orogen, Labrador, Canada: a record of Paleoproterozoic magmatism and deformation. Precambrian Res. 70, 169-190.

Taylor, S. R., McLennan, S.M., 1985. The Continental Crust: its Composition and Evolution, Blackwell Scientific Publications, 312pp.

Taylor, F.C., 1979. Reconnaissance geology of a part of the Precambrian Shield in northeastern Quebec, northern Labrador and Northwestern Territories. Geologica Survey of Canada, Memoir 393: 99pp.

Touret, J.L.R., 1996. LILE-depletion in granulites: Myth or reality? *In* Petrology and Geochemistry of magmatic suites of rocks in the continental and oceanic crusts. A volume dedicated to Professor Jean Michot *Ed.* D Demaiffe: 53-72.

Turcotte, [illegible]L., 1989. Dynamics of recycling *In* Crust/Mantle Recycling at Convergence Zones, Eds. S. Hart and L. Gulen. NATO ASI Ser., 258: 245-257.

Van Kranendonk, M. J., Scott, D. J., 1992. Preliminary report on the geology and structural evolution of the Komaktorvik Zone of the Early Proterozoic Torngat Orogen, Eclipse Harbour area, northern Labrador. *In* Current Research, Part C. Geological Survey of Canada, Paper 92-1C: 59-68.

Van Kranendonk, M. J. and Wardle, R. J., 1996. Burwell domain of the Paleoproterozoic Torngat Orogen: Tilted cross-section of a magmatic arc caught between a rock and a hard place. *In* Precambrian Crustal Evolution in the North Atlantic Region. *Ed.* T. S. Brewer. Spec. Publ. Geol. Soc. Lond. 112: 91-116.

Van Kranendonk, M. J. and Wardle, R. J., (in press). Crustal-scale flexural slip folding during late tectonic amplification of an orogenic boundary pertubation in the Paleoproterozoic Torngat Orogen, NE Canada. Can. J. Earth Sci. (1997)

Van Kranendonk, M.J., Ermanovics, I., 1990. Structural Evolution of the Hudsonian Torngat Orogen in the North River map Area, Labrador: evidence for east-west transpressive collision of Nain and Rae continental blocks. Geosci. Can., 17:283-288.

Wardle, R. J., 1983. Nain-Churchill province cross-section, Nachvak fiord, Northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 83-1: 68-89.

Wardle, R. J., 1984. Nain-Churchill province cross-section, Riviere Baudancourt-Nachvak Lake. *In* Current Research, Newfoundland Department of Mines and Energy, Report 84-11: 1-11.

Wardle, R. J., Van Kranendonk, M.J., Mengel, F., Scott, D., 1992. Geological mapping in the Torngat Orogen, northernmost Labrador: preliminary results. *In* Current Research (1992), Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 92-1: 413-429.

Wyllie, P.J., 1971. The role of water in magma genesis and initiation of diapiric uprise in the mantle. J. Geophys. Res., 76: 1328-1338.

Chapter 2

ARCHEAN CRUSTAL BLOCKS IN NORTHERN LABRADOR: ISOTOPIC AND GEOCHEMICAL CONSTRAINTS ON ARCHEAN CRUSTAL DEVELOPMENT PRIOR TO TORNGAT OROGENESIS

2.1 INTRODUCTION

In the North Atlantic region, Archean cratonic blocks are stitched together by Early Proterozoic collision belts which developed in response to a complex series of tectonic and magmatic processes accompanying continental convergence and collision, including sedimentation, volcanism and arc magmatism. During collision, the Archean and Early Proterozoic rocks were variably intercalated, deformed and migmatized, thereby losing much of their pre-collisional structural and metamorphic characteristics. Prior to Early Proterozoic amalgamation and metamorphism, each of the Archean cratonic blocks were composed of deformed, heterogeneous rock assemblages each preserving complex magmatic, structural and metamorphic histories (e.g., Bridgwater and Schiotte, 1991; Schoitte et al., 1990). Only through geochemical and isotopic characterization of these Archean crustal blocks is it possible to constrain their origin and pre-orogenic histories, including their distribution, geometry and possible genetic relationship between blocks.

The Torngat Orogen (TO) in Northern Labrador, Canada, marks a zone of Early Proterozoic collision between the Archean Nain and Rae Provinces (Fig. 1). Its northern-most segment, the focus of this study, includes a poorly understood, triangular shaped crustal block called the Burwell Domain (Fig. 1). While the area is known to contain Archean gneisses, supracrustal rocks and intrusive Early Proterozoic meta-igneous rocks, the relationship between the Burwell Domain and adjacent Nain or Rae Provinces is poorly constrained. Early Proterozoic meta-igneous rocks of similar age and composition as those in the Burwell Domain are exposed along the western margin of the Nain Province (Fig. 2), suggesting either that the two Archean blocks (i.e., Nain Province and Archean component of Burwell Domain) were juxtaposed prior to

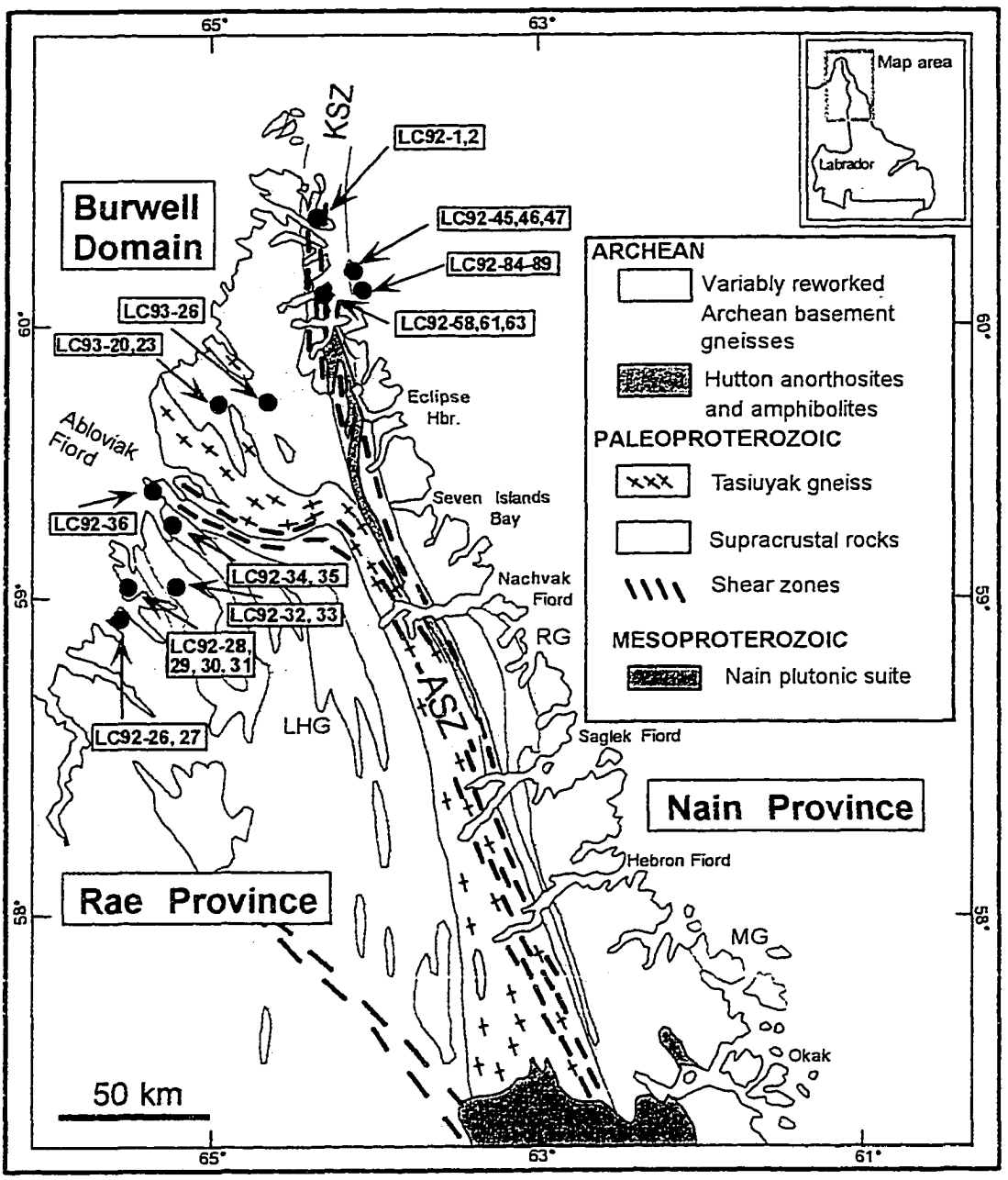


Figure 1. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements, sample locations and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group.

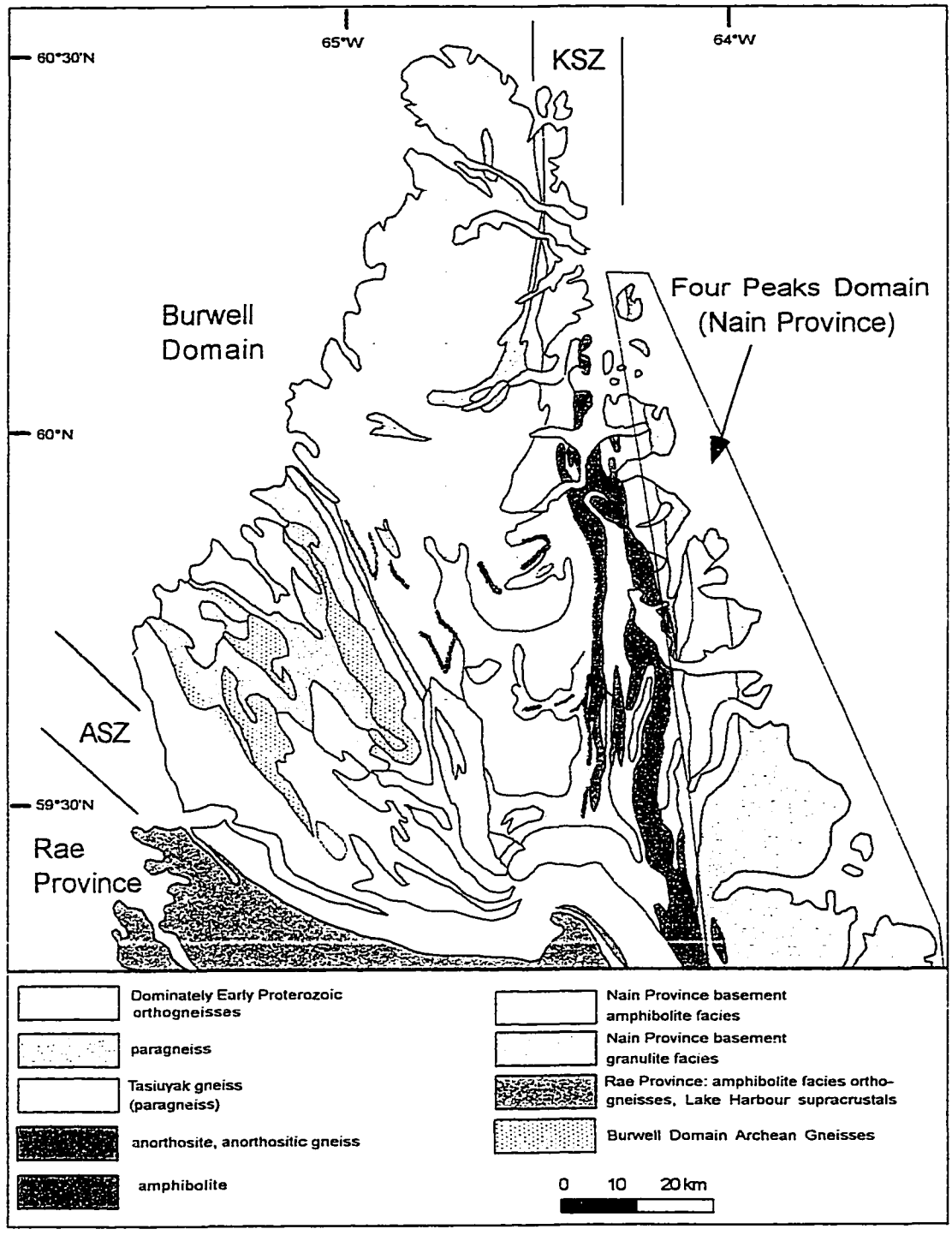


Figure 2. Geological sketch map of the study area in Northern Labrador (for location, see Figure 1). Based on data from Wardle et al. (1993), Van Kranendonk et al. (1994). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively.

Proterozoic magmatism or that Archean gneisses in the Burwell Domain represent a westward extension of the Nain Province.

The aim of this paper is to isotopically and geochemically characterize Archean rocks in the Nain and Rae provinces and in the Burwell Domain in the northern-most segment of the TO. The goals of this work are twofold. The first is to compare geochemical and isotopic data from Archean gneisses in the Burwell Domain with those in adjacent Nain and Rae provinces, in order to evaluate possible genetic links between the two crustal provinces. This is an essential prerequisite for establishing the sequence and geometry of tectonic amalgamation of this region. A second goal of this work is to provide a database for comparisons between Archean rocks in northern Labrador and those occurring in better documented segments of the Nain and Rae cratons farther south (Fig. 3). Within this context, the isotopic and geochemical data from the Northern Labrador Archean gneisses are compared to those used in petrogenetic models describing the generation of Archean high-Al tonalites, trondhjemites and dacites (TTD) occurring worldwide (e.g., Barker and Arth, 1976; Martin, 1993). Finally, the data presented here provide end member crustal compositions necessary for modeling the petrogenesis of the Early Proterozoic magmatic rocks which intrude the region (discussed in Chapter 3).

2.2 GEOLOGICAL OVERVIEW

It is well established that at the latitudes of Okak and Nachvak fjords (Fig. 1), the TO separates strongly deformed and locally migmatized (during Proterozoic convergence) Archean Rae Province gneisses in the west from less deformed (during Proterozoic convergence) Archean Nain Province gneisses in the east (Wardle, 1983, Ermanovics et al., 1989). The presumed suture between the two Archean blocks is marked by the distinctive north-south trending, crustal-scale Abloviak shear zone (ASZ). The ASZ is broadly coincident with the Tasiuyak gneiss (TG; Wardle, 1983), a thick, continuous belt of amphibolite to granulite facies, pelitic to semipelitic paragneisses whose sedimentary protoliths are locally constrained to a 1940-1885 Ma depositional age (Scott

and Machado, 1995). The Tasiuyak gneiss is suggested to represent an accretionary prism comprised of distally derived sedimentary material (Van Kranendonk and Wardle, 1993; Rivers et al., 1996; Theriault and Ermanovics, 1997). It is not clear whether the accretionary prism developed along the margin of the Rae or the Nain Province, however, in the southern TO, ca. 1877 Ma metaigneous rocks intrude both the Rae Province gneisses and Tasiuyak gneiss domain, suggesting that the Tasiuyak gneiss was pinned to the Rae Province prior to 1877 Ma in the southern TO (Bertrand et al., 1993; Theriault and Ermanovics, 1997). Early Proterozoic metaigneous rocks presumed to have formed in a convergent setting via subduction zone processes, occur as intrusive bodies and as tectonically bound sheets along the length of the TO. Combined with structural and metamorphic evidence for crustal scale collision, the presence of these intrusive rocks have led workers to interpret the Proterozoic evolution of the TO within a framework of Phanerozoic-style convergence, involving subduction-related magmatism and continental collision between the Archean Nain and Rae Cratons (see Chapter 1 for a more detailed and fully referenced summary of previous and present models for the development of TO).

2.2.1 The Nain Province

To the south of the present study area the Nain Province preserves some of the oldest continental crust on Earth and therefore many aspects of the geology are well documented, particularly in regions where ca. 3.9-3.7 Ga crust has been identified (e.g., Saglek and Hebron areas of the Saglek block, Figs. 1 and 3). In a model proposed by Schiotte et al. (1990), the Nain Province is described as a heterogeneous tectonic collage of Archean terranes, which was assembled at ca. 2.7-2.8 Ga. Each of the component terranes had a different magmatic, tectonic and metamorphic history prior to assembly. For example, in the Saglek block (Fig. 3), volumetrically minor 3200-3260 Ma Lister gneisses are intercalated with >3500 Ma Uivak gneisses (Schiotte et al., 1991), while in the Hopedale block farther south, ca. 3260-3100 Ma

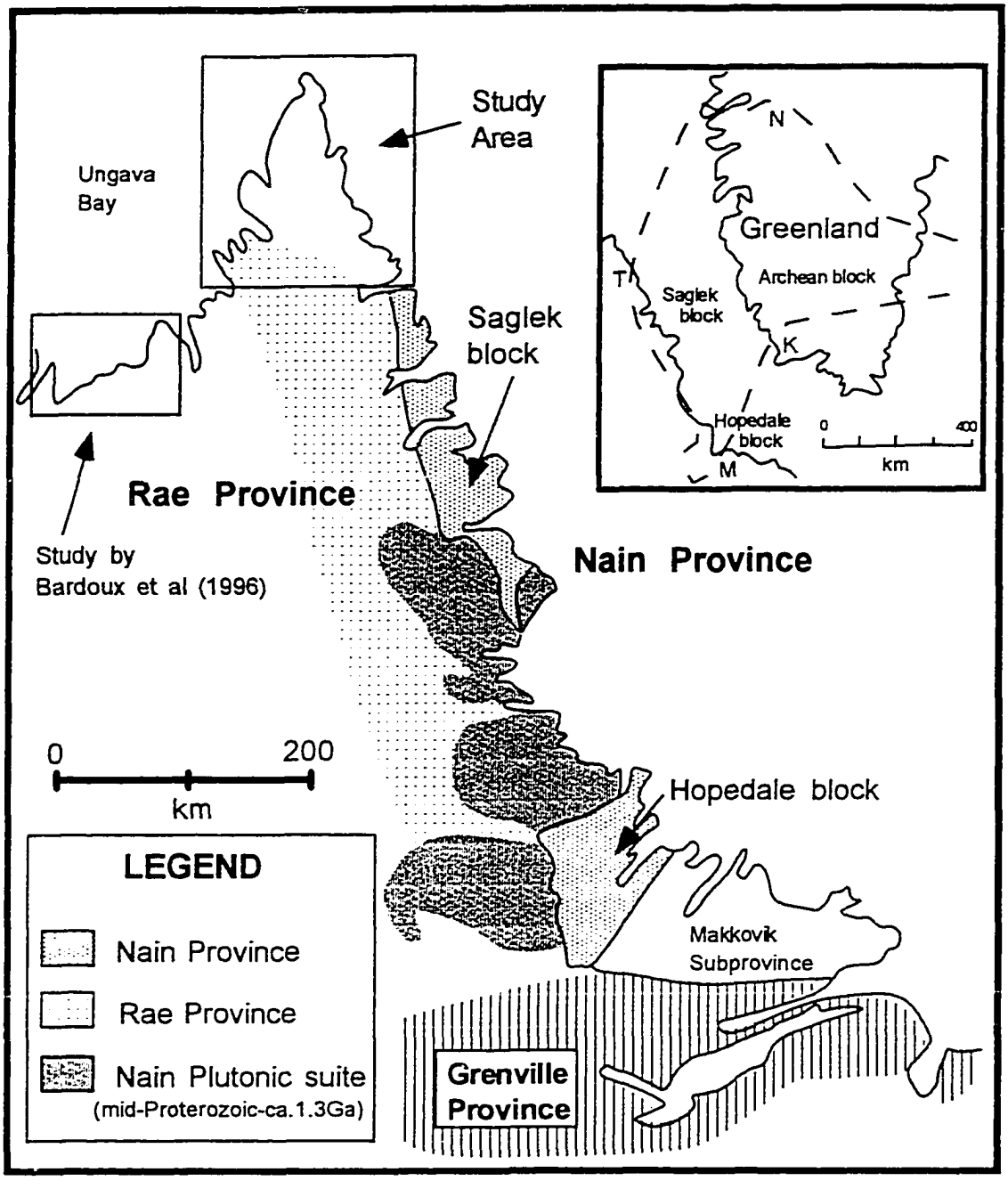


Figure 3. Sketch map of Northern Labrador showing the location of major lithotectonic units discussed in the text, including the Archean Saglek and Hopedale blocks of the Nain Province, and the Paleoproterozoic Makkovik subprovince. The area covered in the study by Bardoux et al. (1996) in Ungava Bay is also indicated. Inset map shows the linkages between the Archean Saglek and Hopedale blocks in Labrador and the Archean block in Southern Greenland. This North Atlantic Archean Craton is bordered to the West and North by the Paleoproterozoic Torngat (T) and Nagssugtoqidian (N) orogens, and to the South by the Makkovikian (M) and Ketalidian (K) orogens. Modified from Bridgwater and Schiotte, 1990 and Collerson et al., 1982.

Maggo gneisses dominate (Loveridge et al., 1987; Schiotte et al., 1991). The similarity in crystallization ages between the Maggo and Lister gneisses suggests a possible genetic relationship between them, which would link the Hopedale and Saglek blocks of the Nain Province to a common magmatic event. The processes that drove crustal assembly are not well established, thus it is not clear how closely the ca. 2.7-2.8 Ga terrane amalgamation resembles better constrained Phanerozoic crust-forming events, or how representative they are of Archean processes on a global scale. It is, however, clear that regional variations in age and crustal history are preserved (pre-2.7-2.8 Ga) in both the Hopedale and Saglek blocks of the Nain Province (for review, see Bridgwater and Schiotte, 1991; Schiotte et al., 1991). In contrast, all rocks from these crustal blocks share the same post-amalgamation history, including a high metamorphic grade which developed in response to collisional events between 2.78-2.74 Ga (Schiotte et al., 1990) and which resulted in a long-lived depletion in U and Rb, causing the characteristically unradiogenic present-day Sr and Pb isotopic signatures for most of the crustal units that comprise the Nain Province.

Another link between Saglek and Hopedale blocks (Fig. 3) is the occurrence of ca. 3.2 Ga gneisses and ca. 3.0 Ga mafic dykes in both areas (Schiotte et al., 1990; Bridgwater and Schiotte, 1991; Ermanovics, 1993). Similar correlations have recently been proposed between the Saglek block and the Four Peaks Domain, an Archean crustal block occurring north of Nachvak Fiord and east of the Komaktorvik shear zone (Fig. 2; Wardle et al., 1993). Along the western margin of the Four Peaks Domain, the Hutton anorthosite and associated amphibolites form a linear belt that extends southward from northernmost Labrador for over 200 km (Fig. 1). Prior to the 1991-1993 field work (for details about field project, see Chapter 1), much of the Four Peaks Domain was considered to be part of the southeast arm of the Rae Province (Taylor, 1979; Hoffman, 1988). However, recent studies have identified granulite facies metamorphic events, Mid- to Late Archean felsic magmatism (e.g., Scott, 1995a) and the occurrence of ca. 2139 Ma mafic dykes in the area (Connelly and Mengel, 1996), all of which correlate with similar events in the Saglek block and

farther south, suggesting the Four Peaks domain is a northern continuation of the Nain Province (Fig. 2; Wardle et al., 1992; Van Kranendonk & Scott, 1992; Bridgwater and Schiotte, 1991).

2.2.2 The Rae Province

Our knowledge concerning crustal evolution in the adjacent Rae Province (Fig. 1) is relatively incomplete. Previously considered as part of the Eastern Churchill Province (Stockwell, 1961), the area was described as having been strongly deformed and metamorphosed during Hudsonian orogenesis (ca. 1.9-1.8 Ga; e.g., Lewry and Collerson, 1990). However, later studies have documented significant lateral variations in the structural and chronological histories of different crustal domains within the Rae Province (e.g., James et al., 1995; Van Kranendonk and Wardle, 1996), and as a result, the area is now viewed as a composite terrane of strongly reworked Archean and Proterozoic rocks.

The segment of Rae Province crust occurring in the northern TO (immediately south of Abloviak fiord; Fig. 1), has not previously been isotopically or geochemically characterized. Wardle (1984) described the eastern occurrences of the Rae Province basement gneisses as broadly similar to the adjacent Nain Province gneisses, based on lithological similarities and on the field observation that both had experienced granulite facies metamorphism prior to Early Proterozoic retrogression. However, unlike the rocks in the Nain Province, the tonalites and amphibolites in the Rae Province are locally intruded by significant volumes of granite (s. s.). Immediately southwest of the current study area, along the southeastern shore of Ungava Bay (Fig. 3), a geochronological study by Bardoux et al. (1996) reports the occurrence of polyphase gray tonalitic gneisses (ca. 2629-2801 Ma) interlayered with amphibolites and intruded by 1853 Ma migmatitic melts (all U-Pb zircon ages). On the basis of the field descriptions in Bardoux et al. (1996), the rocks in the southeastern Ungava Bay section are very similar to those in the area studied here, however, due to the limited isotopic data in this study, rigorous correlations are not yet possible.

2.2.3 Burwell Domain

The Burwell Domain is comprised of an enigmatic collage of Archean gneisses, paragneisses and Early Proterozoic metaigneous rocks (for review, see e.g., Van Kranendonk et al. 1994). The lateral variations of geologic structures and in the relative proportions of lithologies are significant. For example, the northwestern part of the Burwell domain area is largely devoid of supracrustal rocks, whereas in the southwest the Burwell Domain is dominated by granulite facies supracrustal rocks and associated Tasiuyak gneiss. These are intercalated with ca. 1864-1910 Ma Proterozoic plutonic rocks (Scott, 1995a) and with large, tectonically bound blocks of complex gneisses which are known to be Archean on the basis of U-Pb analyses which yield metamorphic ages of 2772 and 1850 Ma (Fig. 2; Connelly and Mengel, 1995).

In the northern segment of the TO, the north-striking ASZ and associated Tasiuyak gneiss are deflected westward and form the southern margin of the Burwell Domain (Fig. 1). At this "big bend", the Komaktorvik shear zone (KSZ), splays off to the north. The KSZ mainly records oblique, sinistral, east-side-up movement which occurred primarily between 1798-1780 Ma (Van Kranendonk and Scott, 1992; Bertrand et al., 1993; see Chapter 1 for more detailed description of chronology of the northern TO). The scale and tectonic significance of the KSZ are not well established, but it has been interpreted as either a tectonic boundary between the western margin of the Nain Province and the eastern edge of the Burwell Domain (e.g. Hoffman, 1988, 1990), or a pre-existing zone of lithospheric weakness within the Nain Province, which was reactivated during crustal readjustments late in Torngat orogenesis (Scott, 1995b; Van Kranendonk and Wardle, 1996). The KSZ has also been suggested to roughly mark the location of an Early Proterozoic back arc basin (Van Kranendonk and Wardle, 1996). Interpretations regarding the significance of the Burwell Domain as a tectonic entity and its linkages with adjacent Rae and Nain Provinces are quite diverse. Hofmann (1988, 1990) viewed the Burwell Domain as a foreign (i.e., unrelated to Nain or Rae provinces) crustal block that was laterally accreted to the western margin of the Nain Province prior to Early Proterozoic magmatism, whereas Taylor (1979) described the Burwell

Domain as the northern continuation of the southeast Rae Province, based on lithological, structural and metamorphic similarities. The remaining option, namely that the Burwell Domain represents the western extension of the Nain Province has been proposed by Korstgaard et al. (1987) on the basis of similarities of aeromagnetic signatures, by Scott and Campbell (1993), Scott (1995b), Rivers et al. (1996) on the basis of the broadly contemporaneous Early Proterozoic magmatism in both areas, and by Connelly and Mengel (1995) on the basis of ca. 2.77 Ga metamorphism that affected the Archean gneisses on both sides of the KSZ.

In this paper we present geochemical and Nd, Sr and Pb isotopic data from the Archean blocks within the Burwell Domain in order to establish the linkages with the adjacent Rae and Nain provinces and to test the above proposals regarding the origin and significance of the Burwell Domain.

2.3 SCOPE OF THIS STUDY

In this study we have sampled thirty-five rocks from the Archean Nain and Rae Provinces and from Archean gneisses within the Burwell domain. Sample locations are shown on Figure 1. The samples were analyzed for major and trace element concentrations, and subsets of these were selected for Sm-Nd (n = 12), Rb-Sr (n = 9) and Pb-Pb (n = 21) isotopic analyses. Most of the outcrops sampled represent complex, heterogeneous gneisses. Therefore, the samples used in this study were taken from relatively thick layers of homogeneous, unmigmatized material, and areas of strong Proterozoic deformation (i.e., showing structural styles related to the ASZ and KSZ) were avoided. In order to facilitate comparisons between lithologies from the Nain and Rae Provinces and Burwell Domain, the rocks have been subdivided on the basis of major element chemistry into 1) tonalites, trondhjemites and granodiorites (or "tonalitic rocks"), 2) granites and quartz monzonites (refered to as "granites"; 3) amphibolites; and 4) anorthosites.

2.4 DESCRIPTION OF SAMPLES

2.4.1 Nain Province

2.4.1.1 Field Description

Basement gneisses in the Nain Province show a large compositional range. The Archean crustal history of the area is complex and includes multiple phases of felsic magmatism (see Bridgwater and Schiotte, 1991). The gneisses are commonly migmatized and compositionally layered on centimeter to meter scales. The majority of the gneisses are tonalitic with layers and pods of amphibolite or gabbro, gabbro-anorthosites, ultramafic rocks, and varying amounts of supracrustal material, including carbonates and rusty-weathering pelites.

In the Nain Province, the anorthosites and juxtaposed amphibolites occur as structually bound linear belts within the KSZ. They define, along with the KSZ, the western margin of the Four Peaks domain (Figs. 1 and 2). Field observations of Early Proterozoic magmatic rocks intruding both the amphibolite-anorthosite belts and Nain Province gneisses suggest that amphibolites and associated anorthosites were juxtaposed with Archean gneisses prior to emplacement of the magmatic rocks at ca. 1.9 Ga (Scott & Machado, 1995), however the sharp tectonic boundaries between structurally complex Nain Province gneisses and less structurally complex amphibolite-anorthosite belts suggest that juxtaposition occurred late during assembly of the Nain Province (i.e., Late Archean or earliest Proterozoic).

2.4.1.2 Petrography

Nain Province tonalitic gneisses typically preserve totally recrystallized granoblastic textures and only a few samples show coarse grained, possibly magmatic textures with domains of incipient recrystallization. Orthopyroxene is common and usually replaced by hornblende along its margins. Late alteration is shown by serpentine along cracks and cleavage planes. Clinopyroxene is rare, and also shows marginal replacement by hornblende. Locally, pyroxene is entirely replaced by hornblende. Biotite is present in all samples, and is commonly kinked and defines a weak foliation.

Oxides are common (ca. 5 vol.%), plagioclase is abundant (20-50 vol.%) and apatite and zircon occur as accessory minerals in most samples (ca. 2-3 vol.% and ca. 1 vol.%, respectively).

Amphibolites are totally recrystallized and strongly foliated. They are composed of clinopyroxene (ca. 20 vol.%), orthopyroxene (10 vol.%), hornblende (> 40 vol %) and plagioclase (ca. 30 vol.%). Adjacent anothosites contain plagioclase (ca. 60-70 vol.%), hornblende with minor orthpyroxene cores (ca. 20 vol.%), biotite (2-3 vol.%) and minor quartz.

2.4.2 Rae Province

2.4.2.1 Field Description

Within the study area (Fig. 1), Rae Province rocks show increasing degrees of deformation and amphibolite facies retrograde metamorphism from the southern-most sampling location, where the rocks are relatively unaffected by Proterozoic deformation, towards the Ahloviak shear zone, where the rocks are more strongly affected by Proterozoic deformation (Goulet, 1992; Rivers et al. 1996). Based on field relationships at the five sampling sites, the Rae Province appears to contain three distinct components: (a) (oldest) amphibolite, which is intruded by (b) tonalitic rocks (tonalites, trondhjemites and granodiorites). These are locally migmatized and intruded by (c) pink granitic rocks.

2.4.2.2 Petrography

Rae Province tonalites have inequigranular, possibly igneous textures preserving grain boundaries that vary from lobate or cuspate to serrate. Biotite is the only mafic phase observed in the tonalitic samples, and ranges from 10 to 30 vol.%. Quartz grains are always strained and make up ca. 30% by volume. Plagioclase comprises 30-40 vol.%, is often strained and is commonly perthitic. Locally, the tonalitic rocks contain up to 15 vol.% K-feldspar.

Rae Province granites preserve igneous textures. Hornblende (ca. 5 vol.%) and biotite (ca. 10 vol.%) are the mafic phases, and Fe-Ti-oxides make up ≤ 5 vol.%. Quartz and K-feldspar are

abundant (ca. 30 vol. % each) and plagioclase constitutes about 15 vol.% of the rock. Sphene is the main accessory phase.

Amphibolites are totally recrystallized, and contain hornblende (ca. 40-45 vol.%; locally with orthopyroxene cores), plagioclase (ca. 40-45 vol.%) and epidote (ca. 10 vol.%).

2.4.3 Archean Gneisses In The Burwell Domain

2.4.3.1 Field Description

Within the western part of the Burwell Domain, Archean gneisses occur as large blocks in the predominantly Proterozoic magmatic terrane (Fig. 2). Intercalated with the Proterozoic magmatic rocks and Archean blocks are abundant enclaves of metasediments (e.g., Van Kranendonk et al., 1994).

Archean gneisses were sampled at two localities in the Burwell Domain (Fig. 1). In the field, the Burwell Archean gneisses were indistinguishable from Nain Province gneisses described above, and showed several phases of migmatization and deformation, in contrast to the adjacent, structurally simple Early Proterozoic lithologies.

2.4.3.2 Petrography

Tonalites in the Buwell Domain contain plagioclase (ca. 35-45 vol.%), K-feldspar (ca. 10 vol.%) and quartz (ca. 35-45 vol.%), and display domains with coarse grained, possibly igneous textures, and domains in which these minerals are recrystallized and strain free. Biotite is the dominant mafic mineral (ca. 5 vol.%), with small amounts of coarse grained hornblende. Garnets occur as large fragmented grains with inclusions of muscovite, K-feldspar and plagioclase. Zircon is the main accessory phase (≤ 1%).

2.5 PRESENTATION OF DATA

2.5.1 Major and Trace Element Geochemistry of the Archean gneisses

Abundances of major elements (recalculated to 100% anhydrous) and trace elements for thirty-five whole-rock samples of Archean orthogneisses occurring in the northernmost segment of the Torngat Orogen are summarized in Tables 1 and 2. Samples are from the Nain Province (n = 20), Rae Province (n = 12) and Burwell Domain (n = 3) and locations are shown in Figure 1. All orthogneisses are calc-alkaline on an AFM diagram (Strekeisen, 1976; Fig. 4A).

2.5.1.1 Nain Province

Tonalitic Rocks

The volumetrically most significant rocks among the Nain Province gneisses in the study area fall in the compositional fields of tonalite, trondhjemite and granodiorite (Fig. 4). They have high Al_2O_3 (14.2-17.7 wt.%; Fig. 5A), moderate to high Sr (> 330 ppm; Fig. 6B), variable Na_20/K_20 ratios (1.1-7.2), low Rb/Sr (0.014-0.18), K/Rb (< 550), and Nb (< 11ppm; Figs. 6 and 7).

On the basis of their silica content they plot in two compositional ranges; felsic (65-74 wt.% SiO_2) and intermediate (SiO_2 = 55-65 wt.%.; Fig. 5, Table 1). The felsic samples are generally homogeneous, however some are migmatized and contain inclusions of mafic layers or pods. Intermediate samples commonly show a small-scale migmatitic layering; the mafic and felsic domains were not separated in the samples analyzed.

Chondrite normalized REE diagrams for the felsic samples (Fig. 8A, B, C), show strong LREE enrichment (La_N = 50-150), less enriched HREE values (Yb_N < 5; La_N/Yb_N = 10-156), and positive Eu anomalies are common (Eu/Eu* > 1). The samples follow three distinct patterns (Fig. 8A, B and C, respectively), which have similar LREE values, but which differ in the magnitudes of Eu anomalies and HREE depletions. Intermediate samples show a comparable LREE enrichment (La_N = 100-500; Fig. 8D), but have higher Y and HREE concentrations (Yb_N = ca. 8.5-15; La_N/Yb_N = 10-38) than the more felsic samples and have negative Eu anomalies (Eu/Eu* = 0.72-0.80).

TABLE 1 - Major element analyses of Archean gneisses from the Nain Province, Rae Province and Burwell Domain, Torngat Orogen, Northern Labrador

	NAIN PROVINCE									
	Tonalitic rocks									
(Wt.%)	LC92-001	LC92-045	LC92-077	LC92-081	LC92-085	LC92-086	LC92-087	LC92-088	LC92-089	LC92-093A
SiO2	70.11	66.94	65.86	70.44	73.08	71.41	73.33	71.40	71.09	73.17
TiO2	0.19	0.68	0.95	0.30	0.16	0.34	0.25	0.27	0.47	0.17
Al2O3	15.87	15.54	16.24	15.83	15.37	15.53	14.95	15.75	14.22	15.16
Fe2O3(T)	2.34	4.67	3.91	3.31	1.30	2.62	1.81	2.02	3.70	1.51
MnO	0.03	0.05	0.03	0.05	0.01	0.01	0.02	0.02	0.05	0.03
MgO	1.30	1.90	2.76	1.04	0.58	0.80	0.56	0.62	1.47	0.39
CaO	3.41	5.26	4.60	3.59	3.39	3.98	3.25	3.31	2.78	1.79
Na2O	4.78	4.16	4.41	4.61	3.89	4.03	4.01	3.94	3.20	4.32
K2O	1.88	0.58	1.03	0.74	2.20	1.20	1.74	2.59	2.92	3.39
P2O5	0.07	0.20	0.20	0.09	0.02	0.06	0.09	0.08	0.09	0.08
H2O(T)	0.20	0.30	0.50	0.30	0.30	0.50	0.30	0.40	0.30	0.30
CO2(T)	0.10	0.50	0.10	0.10	0.30	0.20	0.30	0.40	0.20	0.20
Sum	98.27	98.44	99.75	99.80	100.17	99.14	99.69	99.02	99.88	99.63

	NAIN PROVINCE									
	Intermediate			Amphibolites				Anorthosites		
	LC92-002	LC92-047B	LC92-084	LC92-061	LC92-063	LC92-092A	LC92-092C	LC92-019	LC92-058	LC92-071
SiO2	59.19	55.93	64.60	48.02	48.20	47.49	47.81	54.34	53.07	53.92
TiO2	0.64	0.82	0.70	1.01	1.01	0.65	0.90	0.13	0.47	0.63
Al2O3	17.94	17.36	16.10	14.80	13.49	16.27	15.03	23.60	22.90	18.17
Fe2O3(T)	5.47	8.22	5.33	12.38	14.98	12.05	13.53	3.92	5.25	9.29
MnO	0.09	0.09	0.05	0.21	0.25	0.17	0.20	0.05	0.05	0.14
MgO	2.58	4.44	2.54	7.17	6.58	9.54	8.85	3.43	3.44	5.58
CaO	5.35	7.83	4.82	13.59	13.59	11.75	11.53	9.28	9.41	8.03
Na2O	5.68	4.26	4.02	2.52	1.69	1.81	2.00	4.82	4.84	3.79
K2O	2.70	0.73	1.63	0.23	0.14	0.23	0.08	0.41	0.49	0.36
P2O5	0.36	0.32	0.20	0.08	0.08	0.05	0.06	0.02	0.06	0.08
H2O(T)	0.40	0.60	0.50	0.60	0.70	0.70	0.90	0.80	0.40	0.90
CO2(T)	0.10	0.50	0.50	0.10	0.20	0.30	0.50	0.20	0.10	0.10
Sum	98.67	98.52	99.38	99.54	100.82	99.60	99.77	99.56	99.31	100.14

Major element data by X-ray fluorescence except Fe2O3 by ICP.
Oxides normalized to 100% (anhydrous). Actual analysis totals are given in "Sum" column.
Total iron reported as Fe2O3.

TABLE 1 - (continued)

	RAE PROVINCE									
	Tonalitic rocks					Granitic rocks				
(Wt. %)	LC92-026	LC92-032A	LC92-032B	LC92-033	LC92-034	LC92-029A	LC92-029B	LC92-030	LC92-031A	LC92-031B
SiO2	76.89	73.26	71.31	73.12	67.03	69.66	63.65	70.46	65.77	65.02
TiO2	0.35	0.23	0.27	0.18	0.40	0.64	1.19	0.32	0.83	1.06
Al2O3	12.47	14.81	15.83	14.58	17.24	14.40	15.26	15.12	15.84	15.13
Fe2O3(T)	1.89	1.71	1.80	1.31	2.82	4.23	6.68	2.32	4.98	6.21
MnO	0.04	0.01	0.01	0.01	0.03	0.06	0.10	0.03	0.08	0.10
MgO	1.51	1.11	1.18	0.85	1.75	0.64	1.25	0.75	0.84	0.99
CaO	2.72	2.53	2.52	1.16	4.34	2.29	3.78	2.04	3.37	3.32
Na2O	2.69	4.74	4.91	3.22	4.54	3.02	3.39	2.92	3.69	3.41
K2O	1.41	1.53	2.07	5.51	1.73	4.83	4.13	5.83	4.27	4.36
P2O5	0.03	0.07	0.09	0.06	0.11	0.23	0.56	0.22	0.34	0.41
H2O(T)	0.50	0.50	0.50	0.50	0.50	0.30	0.40	0.40	0.30	0.40
CO2(T)	0.10	0.20	0.20	0.20	0.10	0.20	0.20	0.30	0.20	0.20
Sum	100.27	99.24	99.84	99.43	99.21	99.34	100.23	99.21	100.35	99.82

	RAE PROV.		BURWELL DOMAIN		
	Amphibolites		Tonalitic rocks		Anor.
(Wt. %)	LC92-027	LC92-028	LC93-023	LC93-026	LC93-020
SiO2	49.56	53.09	72.42	70.79	49.32
TiO2	0.71	2.94	0.24	0.36	0.56
Al2O3	14.37	12.10	15.31	15.69	24.51
Fe2O3(T)	13.68	16.70	1.71	2.61	5.83
MnO	0.19	0.22	0.02	0.04	0.07
MgO	9.39	3.10	0.71	1.08	3.39
CaO	9.91	6.80	2.61	4.23	12.36
Na2O	1.69	3.00	4.13	3.92	3.21
K2O	0.45	1.46	2.75	1.18	0.65
P2O5	0.05	0.60	0.10	0.10	0.10
H2O(T)	1.00	0.80	0.40	0.50	1.10
CO2(T)	0.10	0.20	0.10	0.40	0.30
Sum	100.88	100.02	99.28	99.45	99.55

Major element data by X-ray fluorescence except Fe2O3 by ICP.
Oxides normalized to 100% (anhydrous). Actual analysis totals are given in "Sum" column.
Total iron reported as Fe2O3.
"Anor." = Anorthosite

TABLE 2 - Trace element analyses from Archean gneisses, Torngat Orogen, Northern Labrador

	NAIN PROVINCE												
	Tonalitic rocks										Intermediate rocks		
	LC92-001	LC92-045	LC92-077	LC92-081	LC92-085	LC92-086	LC92-087	LC92-088	LC92-089	LC92-093A	LC92-002	LC92-047B	LC92-084
Ba	580	360	810	630	950	740	590	980	2200	810	1800	370	630
Co	11	14				9		8			21	29	
Cr	25	29				12		<10			41	110	
Cs			0.15	0.06	0.03		0.08		0.09	0.11			0.13
Cu	<10	12				13		15			38	46	
Ga	19	18	18	16	16	18	18	19	17	20	23	23	20
Hf	2.4	5.1	6.6	3.1	4.1	4.1	5.6	4.2	7.3	2.8	5.7	4.1	4.4
In	0.03	0.07	<0.05	<0.05	<0.05	<0.02	<0.05	0.03	<0.05	<0.05	0.12	0.07	0.10
Nb	2.0	4.9	2.4	1.6	1.3	1.1	3.3	2.6	6.6	3.3	12.0	6.6	6.1
Ni	42	23				21		12			38	84	
Pb	31	<20	6	7	12	<20	12	<20	13	25	26	<20	10
Rb	31	<10	29	15	48	17	36	44	56	83	23	<10	47
Sc	4.2	6.5				2		2.4			8.3	18	
Sr	500	710	520	520	380	520	380	450	440	460	1600	700	330
Ta	0.23	0.14	<0.2	<0.2	<0.2	0.07	<0.2	0.08	0.23	<0.2	0.33	0.18	0.25
Th	0.38	0.07	0.05	0.09	0.06	0.15	0.44	0.53	1.30	7.70	3.30	0.08	0.69
Tl	0.12	<0.02	0.08	0.03	0.17	0.07	0.11	0.17	0.22	0.36	0.09	<0.02	0.17
U	0.28	0.11	0.09	0.07	0.11	0.14	0.28	0.20	0.32	0.46	0.25	0.13	0.15
V	17	55				20		10			67	110	
Y	3.9	10	2.8	6.0	1.3	0.9	4.5	6.3	4.4	5.5	26	31	16
Zn	83	49				36		33			96	96	
Zr	94	220	340	160	170	160	260	170	340	110	270	180	200
La	16	23	12	10	16	23	23	23	31	28	100	35	25
Ce	28	45	23	16	23	32	39	39	52	50	200	92	54
Pr			2.4	1.3	1.7		3.8		4.8	4.9			6.1
Nd	10.0	22.0	8.7	4.2	4.7	8.6	13.0	14.0	16.0	16.0	96.0	59.0	24.0
Sm	1.7	3.8	1.4	0.6	0.5	0.8	1.9	2.1	2.0	2.4	16.0	11.0	4.6
Eu	0.66	1.40	0.86	0.51	0.71	1.10	0.98	1.30	1.20	0.61	3.50	2.30	1.00
Gd	1.10	2.80	0.96	0.74	0.28	0.42	1.40	1.70	1.20	1.50	9.50	7.80	3.70
Tb			0.12	0.13	0.03		0.17		0.15	0.21			0.54
Dy	0.77	1.80	0.52	0.91	0.18	0.15	0.89	1.10	0.81	1.00	5.10	5.60	2.90
Ho	0.14	0.34	0.10	0.22	0.04	0.02	0.16	0.23	0.15	0.18	0.85	1.10	0.58
Er	0.37	0.88	0.25	0.66	0.14	0.08	0.43	0.62	0.44	0.46	2.00	2.80	1.60
Tm	0.06	0.13	0.04	0.10	0.03	<0.02	0.06	0.10	0.07	0.05	0.31	0.42	0.22
Yb	0.32	0.78	0.24	0.72	0.22	0.10	0.47	0.58	0.59	0.39	1.70	2.40	1.40
Lu			0.04	0.10	0.04		0.07		0.10	0.05			0.20

Trace element data by ICP-MS (see Appendix 1 for analytical details).

TABLE 2 (continued)

	NAIN PROVINCE							RAE PROVINCE				
	Amphibolites				Anorthosites			Tonalitic rocks				
	LC92-061	LC92-063	LC92-092A	LC92-092C	LC92-019	LC92-058	LC92-071	LC92-026	LC92-032A	LC92-032B	LC92-033	LC92-034
Ba	70	70	50	60	210	290	240	210	430	720	2400	450
Co	59		63			27						
Cr	270		250			59						
Cs		0.03		0.05	0.09		0.03	0.80	0.30	0.30	0.33	0.16
Cu	29		<10			<10						
Ga	16	18	14	16	18	21	18	13	15	16	13	20
Hf	1.9	1.9	1.1	1.3	0.3	1.2	1.1	1.8	3.1	3.8	2.5	3.7
In	0.10	0.09	0.05	0.06	0.07	<0.02	<0.05	0.07	0.07	0.07	0.07	<0.05
Nb	2.6	3.0	1.8	2.0	0.2	2.7	1.9	3.4	2.6	2.4	2.1	4.2
Ni	150		240			52						
Pb	<20	11	<20	2	2	<20	3	15	10	12	20	7
Rb	11	13	11	12	19	14	11	70	58	69	130	80
Sc	40		33			7.5						
Sr	150	120	98	92	690	750	580	120	340	370	330	540
Ta	0.14	0.31	0.19	<0.2	<0.2	0.09	<0.2	0.33	0.36	<0.2	<0.2	0.31
Th	0.11	0.11	0.24	0.11	0.16	0.18	0.07	3.80	4.50	5.40	7.90	0.59
Tl	<0.02	<0.02	0.03	<0.02	0.03	<0.02	<0.02	0.35	0.25	0.33	0.56	0.35
U	0.07	0.09	0.08	0.03	0.03	0.07	0.08	0.75	0.28	0.25	0.16	0.13
V	260		190			43						
Y	24	25	15	21	1.5	5.9	8.8	5.7	2.4	2.6	1.5	3.9
Zn	94		92			58						
Zr	66	61	39	47	13	47	47	85	130	170	120	170
La	4.8	4.7	2.7	2.7	2.6	8.7	6.9	16	16	22	19	23
Ce	12	14	6.7	8.2	4.8	18	16	36	31	39	34	42
Pr		1.9		1.2	0.6		1.9	3.8	3.0	3.8	3.2	4.1
Nd	8.9	10.0	4.9	6.5	2.0	9.2	8.0	14.0	10.0	13.0	10.0	13.0
Sm	2.6	3.2	1.5	2.3	0.4	1.8	1.8	2.5	1.5	1.9	1.4	2.0
Eu	0.97	0.98	0.58	0.85	0.45	1.10	1.00	0.58	0.47	0.47	0.41	0.82
Gd	3.50	4.10	1.90	3.10	0.31	1.40	1.60	1.80	0.87	1.00	0.73	1.30
Tb		0.73		0.59	0.04		0.26	0.24	0.10	0.12	0.07	0.16
Dy	3.80	4.30	2.30	3.60	0.25	1.10	1.60	1.20	0.45	0.53	0.31	0.77
Ho	0.81	0.91	0.50	0.80	0.05	0.20	0.32	0.20	0.07	0.09	0.05	0.15
Er	2.30	2.60	1.50	2.40	0.15	0.52	0.91	0.49	0.18	0.23	0.12	0.34
Tm	0.37	0.38	0.23	0.37	0.03	0.07	0.14	0.07	0.03	0.03	0.00	0.05
Yb	2.30	2.70	1.50	2.50	0.18	0.46	0.98	0.47	0.18	0.17	0.10	0.33
Lu		0.38		0.35	0.03		0.14	0.06	0.03	0.03	0.03	0.05

Trace element data by ICP-MS (see Appendix 1 for analytical details).

TABLE 2 - (continued)

	RAE PROVINCE							BURWELL DOM.		
	Granitic rocks					Amph.		Ton. rocks		An.
	LC92-029A	LC92-029B	LC92-030	LC92-031A	LC92-031B	LC92-027	LC92-028	LC93-023	LC93-026	LC93-020
Ba	1800	1700	2000	2200	2300	70	700	630	330	140
Co										
Cr										
Cs	0.21	0.14	0.24	0.12	0.12	<0.02	0.03	0.32	0.82	0.05
Cu										
Ga	18	20	17	20	20	13	18	20	19	20
Hf	7.1	9.0	5.4	7.5	9.1	1.2	3.4	3.9	4.6	0.9
In	0.06	0.08	<0.05	0.05	0.13	0.13	0.13	<0.05	<0.05	<0.05
Nb	10.0	17.0	5.0	12.0	15.0	1.5	7.9	4.0	4.1	2.7
Ni										
Pb	22	17	25	24	23	2	11	20	7.6	5.5
Rb	140	110	160	100	110	13	28	78	40	15
Sc										
Sr	340	400	380	470	450	72	290	360	290	360
Ta	0.40	0.66	0.21	0.65	0.64	0.20	0.51	0.32	<0.2	1.30
Th	5.70	7.90	55.0	3.30	3.40	0.11	2.60	2.70	3.60	0.67
Tl	0.57	0.52	0.74	0.46	0.48	0.03	0.16	0.47	0.20	0.04
U	0.39	0.45	0.48	0.48	0.55	0.06	0.43	0.40	0.69	0.27
V										
Y	29	35	18	28	31	18	30	4	7.1	6.7
Zn										
Zr	350	440	260	380	500	37	150	150	190	32
La	64	73	<100	61	64	2	24	21	17	4.9
Ce	140	160	440	130	130	5.9	53	40	33	11
Pr	17.0	20.0	41.0	16.0	17.0	0.9	6.5	3.9	3.5	1.3
Nd	67.0	80.0	136	63.0	68.0	4.7	27.0	15.0	14.0	5.9
Sm	12.0	15.0	16.0	12.0	13.0	1.7	6.7	2.2	2.4	1.3
Eu	2.20	2.60	1.90	3.00	3.00	0.70	2.20	0.74	0.64	0.57
Gd	8.50	11.0	7.80	8.00	8.90	2.50	6.50	1.50	1.90	1.30
Tb	1.20	1.40	0.89	1.10	1.20	0.47	1.00	0.18	0.28	0.21
Dy	5.90	7.00	4.00	5.60	6.30	2.80	5.40	0.80	1.30	1.20
Ho	1.10	1.40	0.68	1.10	1.20	0.65	1.10	0.13	0.25	0.24
Er	2.80	3.40	1.60	2.70	3.00	1.90	2.90	0.32	0.67	0.67
Tm	0.41	0.46	0.24	0.40	0.45	0.27	0.44	0.05	0.11	0.11
Yb	2.60	2.90	1.40	2.70	2.90	2.00	3.00	0.31	0.61	0.64
Lu	0.33	0.38	0.19	0.36	0.42	0.31	0.41	0.05	0.10	0.10

Trace element data by ICP-MS (see Appendix 1 for analytical details).
"Amph." = Amphibolites; "Ton." = Tonalitic; "An." = Anorthosite

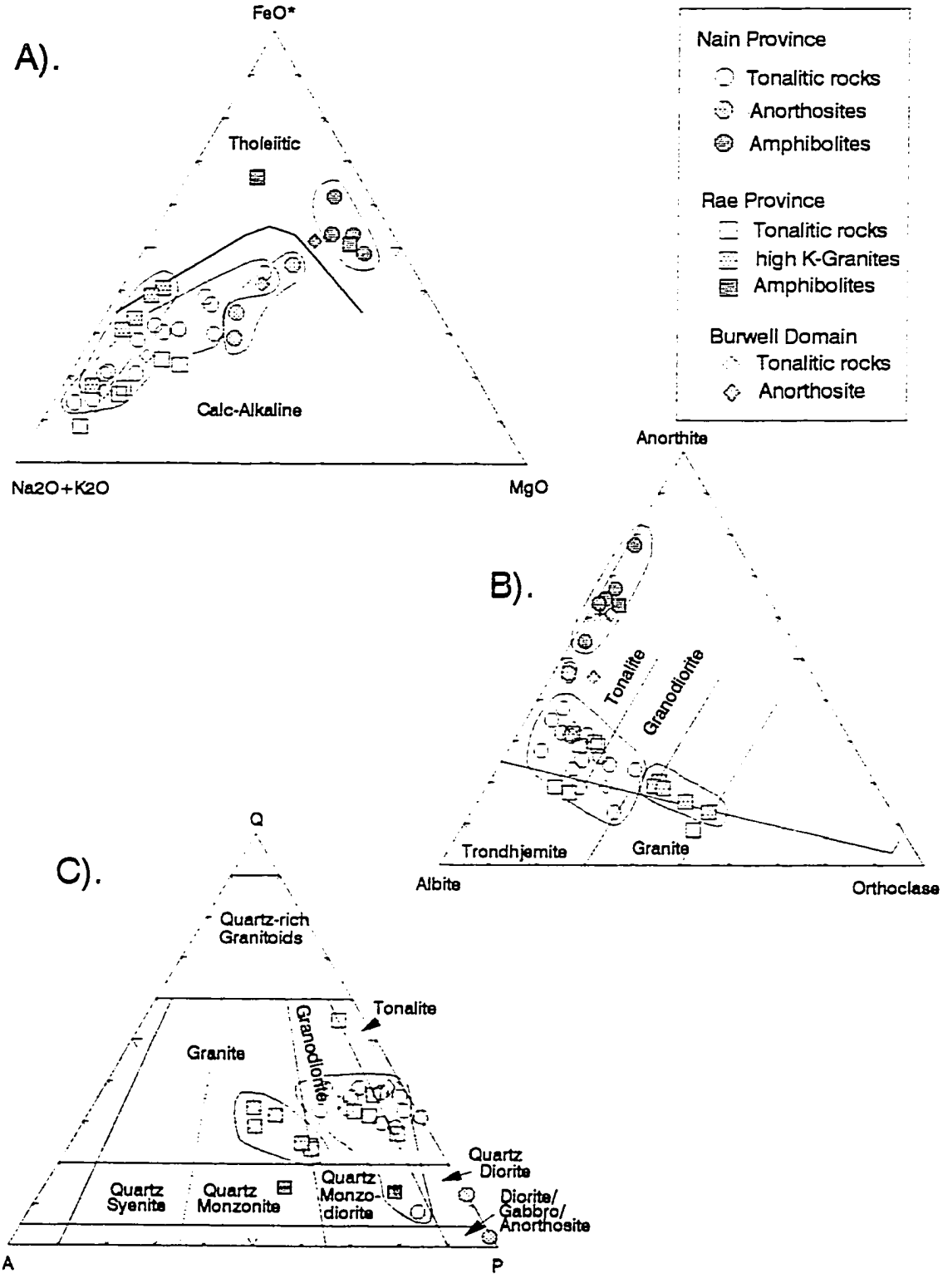


Figure 4. Compositional classification of Archean rocks from the Torngat Orogen, Northern Labrador. **(A)** - AFM diagram (Irvine & Baragar, 1971). FeO* = total Fe calculated as FeO. **(B)** - CIPW normative Anorthite-Albite-Orthoclase diagram (fields from Barker, 1979). **(C)** - CIPW normative quartz-alkali-feldspar-plagioclase diagram from Le Maitre (1989), after Streckeisen (1976).

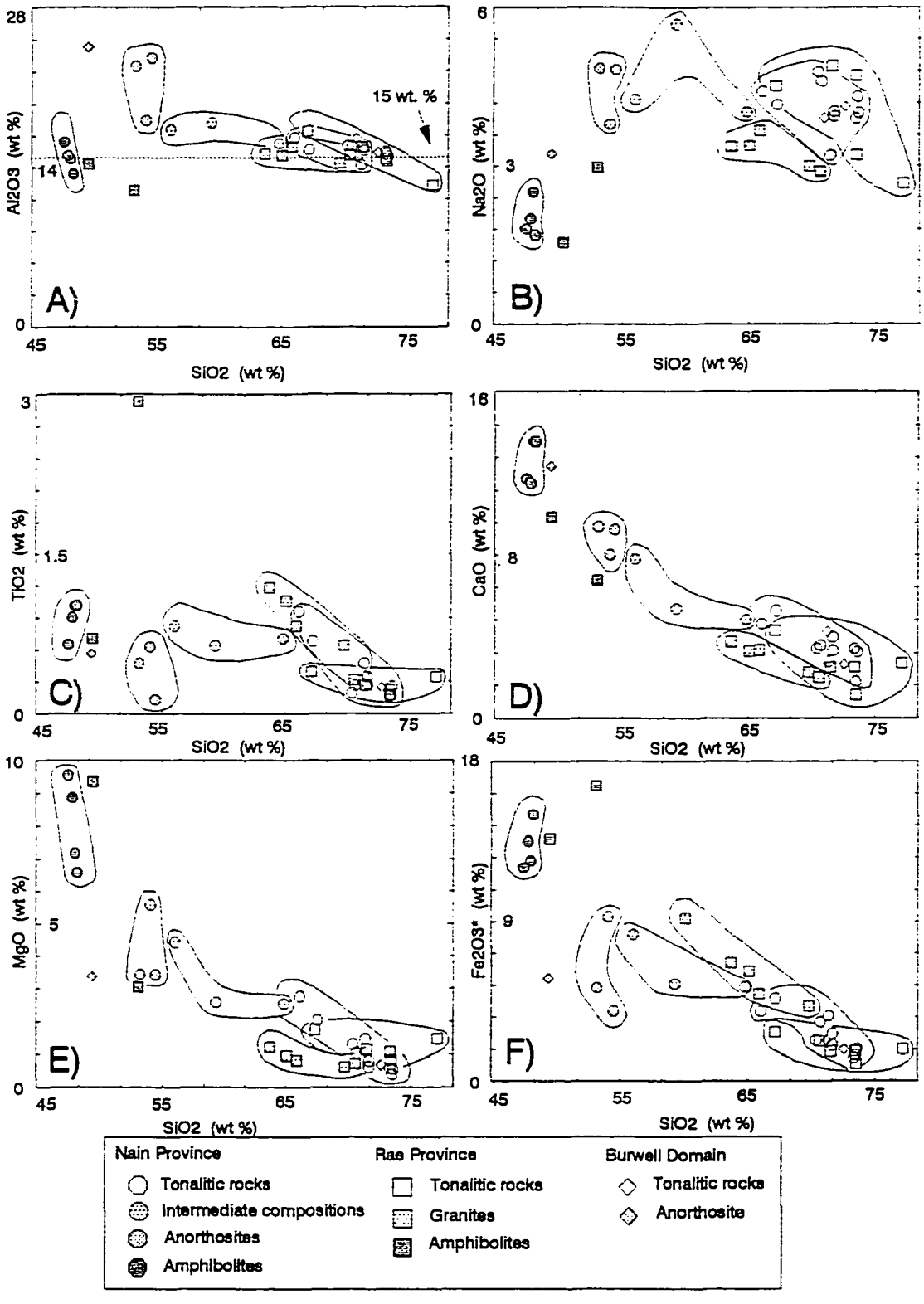


Figure 5. Selected major elements vs. SiO_2. for Archean rocks from the Torngat Orogen, Northern Labrador. Total Fe is given as Fe_2O_3.

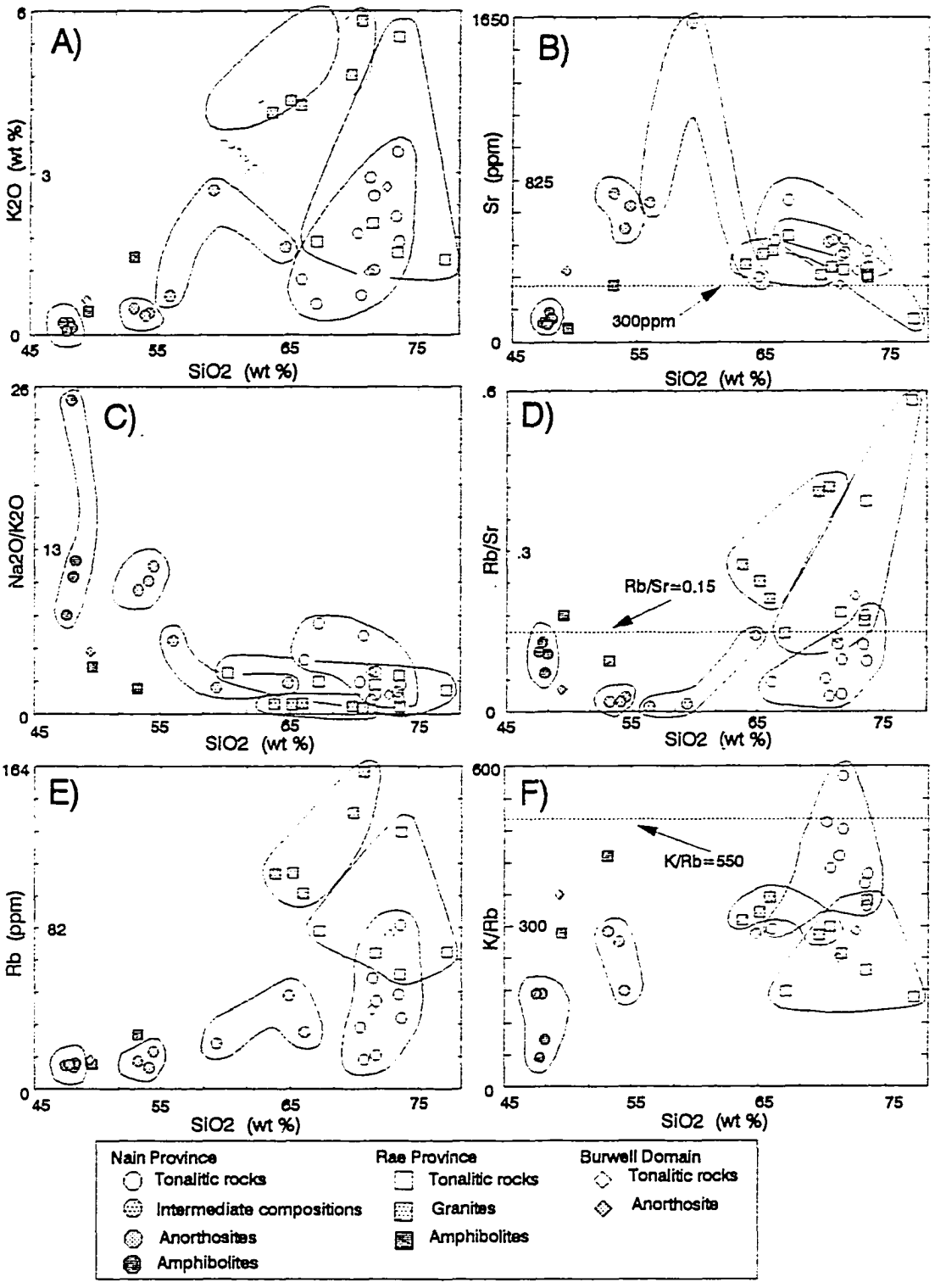


Figure 6. Selected major and trace elements vs. SiO_2 for Archean rocks from the Torngat Orogen, Northern Labrador.

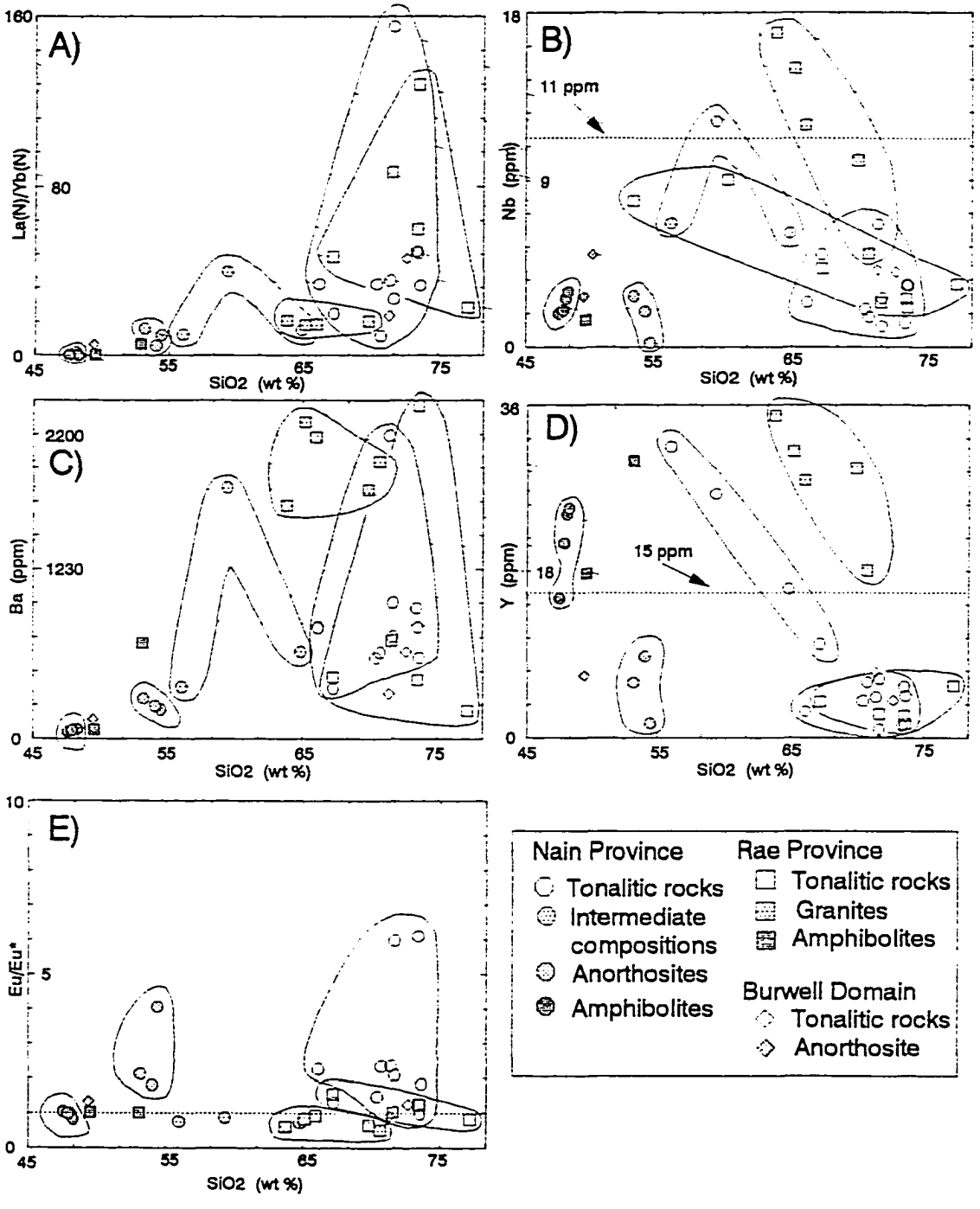


Figure 7. Selected trace and rare earth elements vs. SiO_2 for Archean rocks from the Torngat Orogen, Northern Labrador.

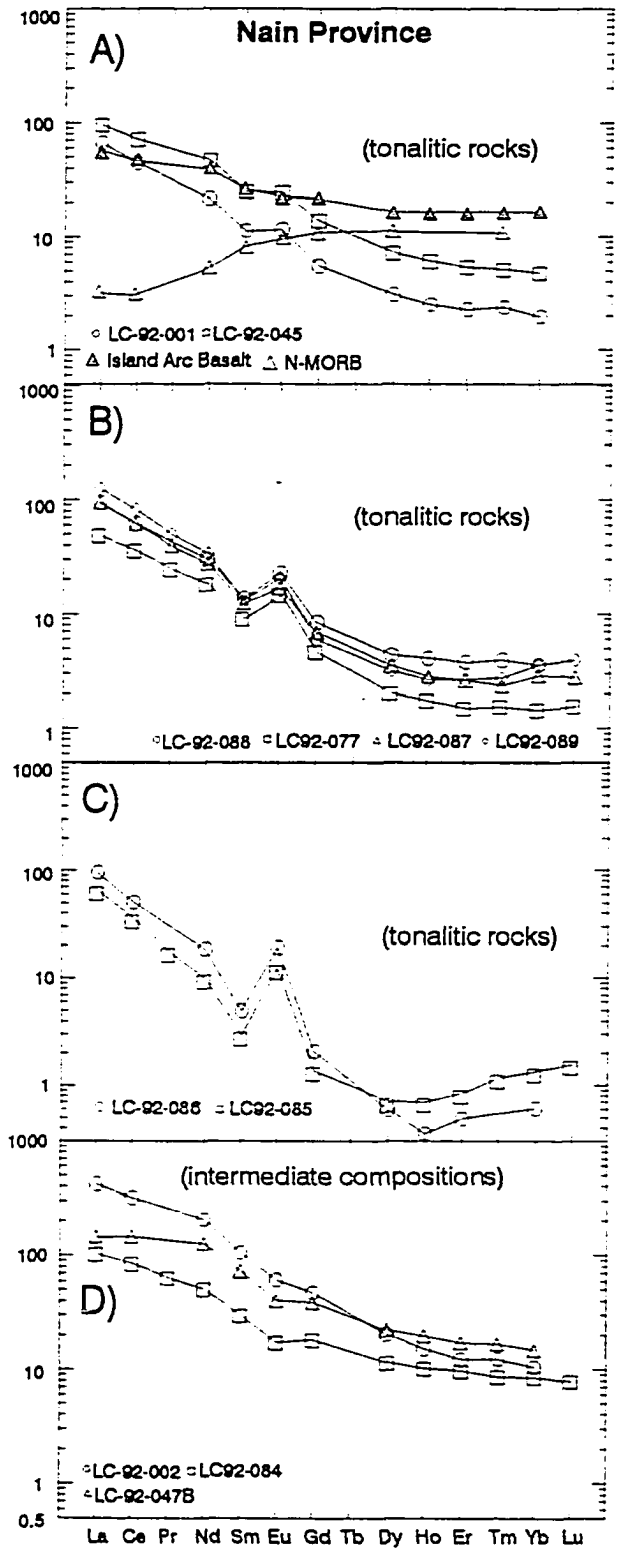


Figure 8. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). **(A)+(B)+(C)** - Tonalitic gneisses from the Nain Province. Samples distributed between diagrams based on REE patterns. **(D)** - Gneisses of intermediate compositions from the Nain Province.

In a MORB normalized spider diagram the felsic samples are subdivided on the basis of their REE patterns (Figs. 9A-C). The samples are enriched in Rb through Nd, with the exception of Th, which is variably depleted relative to adjacent elements U and Sr. For all samples, Zr is slightly enriched relative to adjacent elements, and Ti, Y and Yb are depleted relative to MORB. Intermediate samples (Fig. 10A) show similar MORB-normalized patterns for the elements Rb through Nd, but have higher concentrations of Sm through Yb. They show similar Th depletion and Zr enrichment as the feisic samples.

Amphibolites

The amphibolites are tholeiitic (Fig. 4) with 47.3-48.6 wt.% SiO_2 and 13.6-16.2 wt.% Al_2O_3 (Fig. 5). CaO, Fe and Mg are high (Fe_2O_3 = 12 - 15.1 wt.%; MgO = 8.8 - 9.5 wt.%), Na_2O/K_2O is high, but variable (7.8-25) and Rb/Sr is low (Figs. 6C and D).

Amphibolites yield flat REE patterns at ca. 10 - 20 times chondrite (Fig. 11A) with no Eu anomalies (Eu/Eu* = 0.83-1.05). They are enriched in Rb, Ba, Pb (± K) relative to MORB (Fig. 10B), but less so than adjacent tonalitic gneisses. The amphibolites have MORB values for Sr through Nb with a very minor depletion in Ti, Y and Yb.

Anorthosites

The anorthosites have 52.6-54.1 wt.% SiO_2 and 18.2-23.5 wt.% Al_2O_3 (Fig.5A), Sr is high (580-750 ppm), Rb/Sr is low (< 0.03), and Na_2O/K_2O is high (10-12; Fig. 6). Chondrite-normalized REE patterns show moderate fractionation between light and heavy REEs, with some LREE enrichment (La_N = 10-40; Yb_N = 1-6), and a positive Eu anomaly (Eu/Eu* = 1.77-3.96; Fig. 11B). On the MORB-normalized spider diagram (Fig. 10C), anorthosites are enriched in Rb, Ba, Pb, K, and particularly in Sr. U and Th have MORB-like values, and Zr, Ti, Y and Yb are depleted relative to MORB. Anomalously high Sr and Eu supports formation of the anorthosites as cumulates by fractionational crystallization of plagioclase.

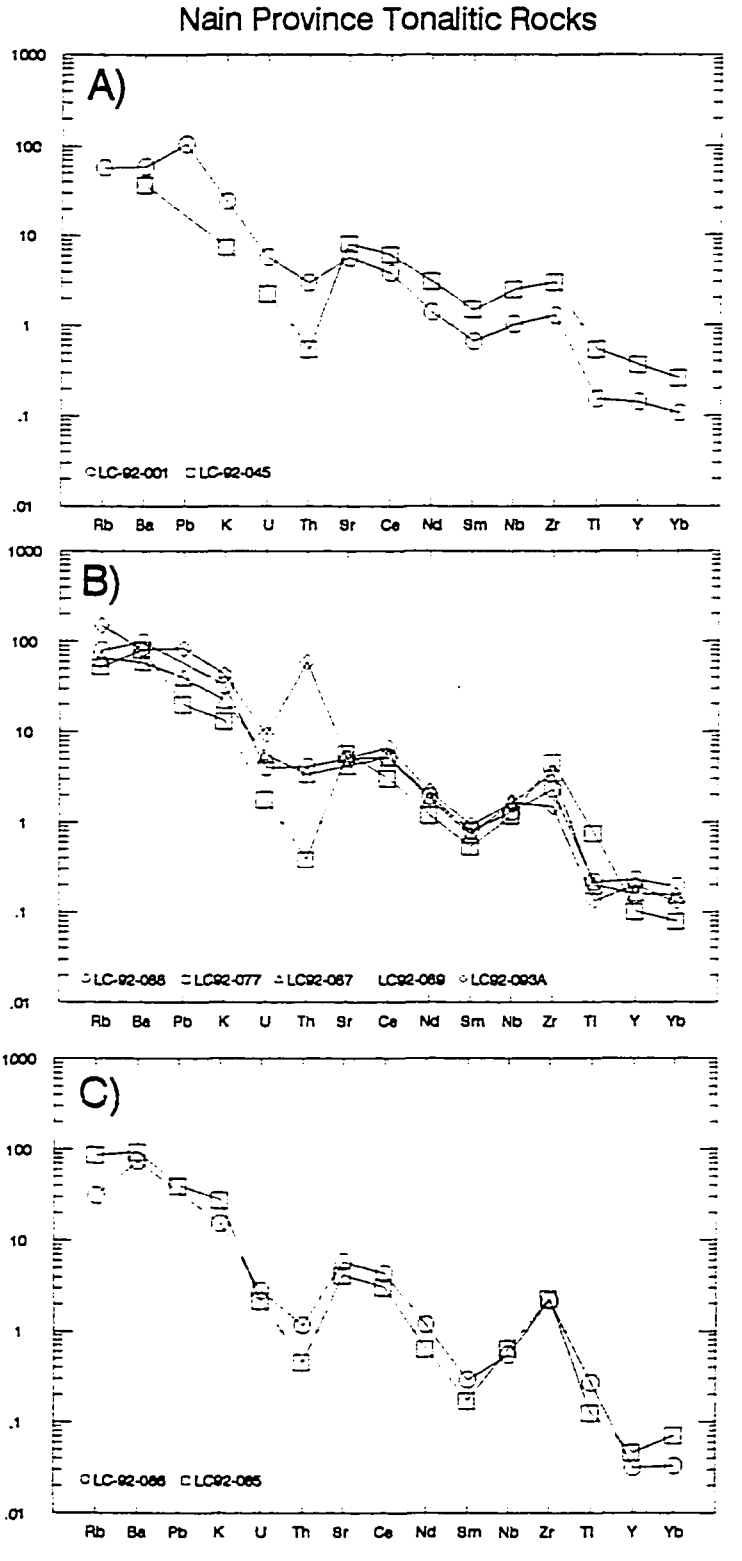


Figure 9. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). **(A)+(B)+(C)** - Tonalitic rocks from the Nain Province distributed between diagrams according to REE patterns (see Figure 8A, B and C).

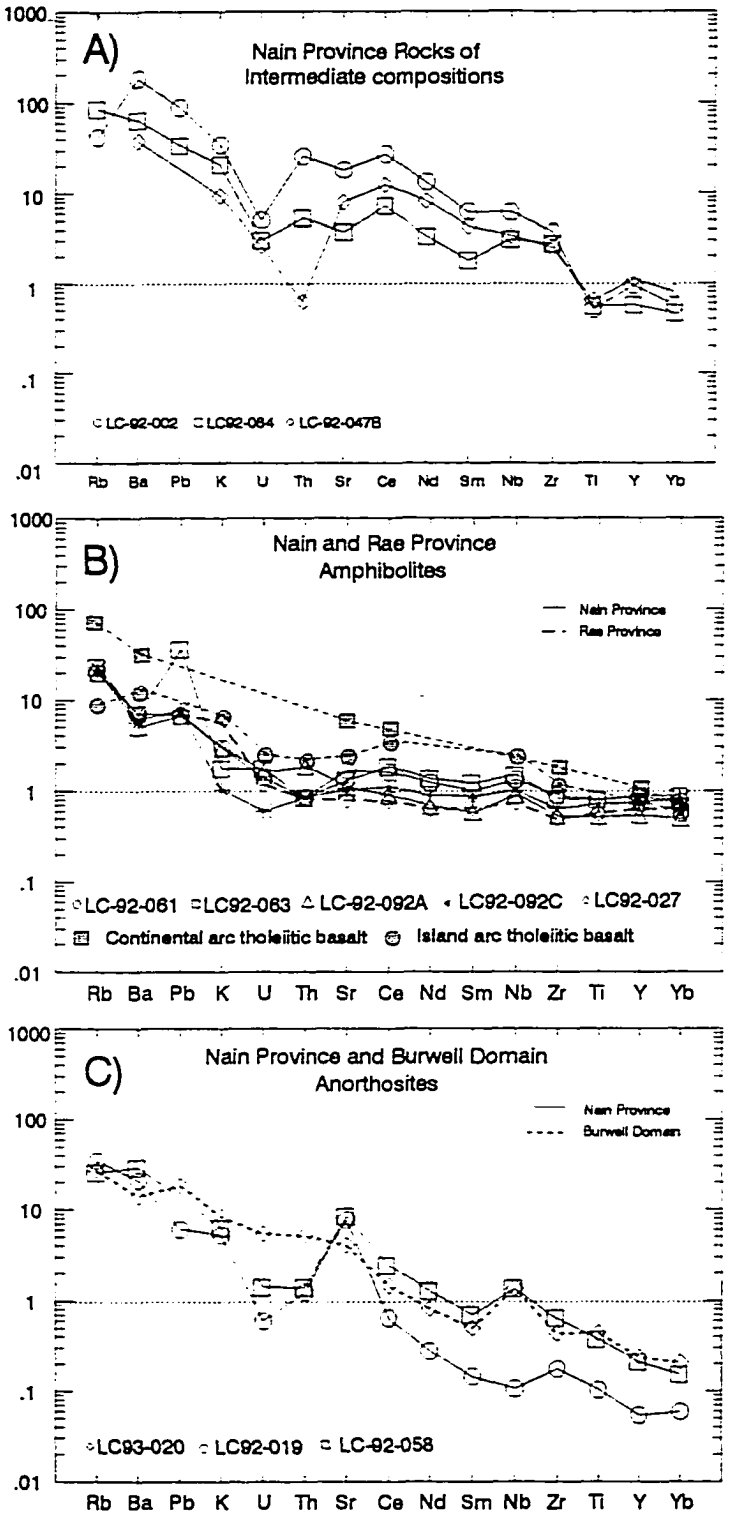


Figure 10. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). **(A)** - Tonalitic gneisses of intermediate composition from the Nain Province. **(B)** - Amphibolites from the Nain and Rae provinces. **(C)** - Anorthosites from the Nain Province and Burwell Domain. Data for continental arc tholeiitic basalt from Ewart (1982), and for island arc tholeiitic basalt from Sun (1980).

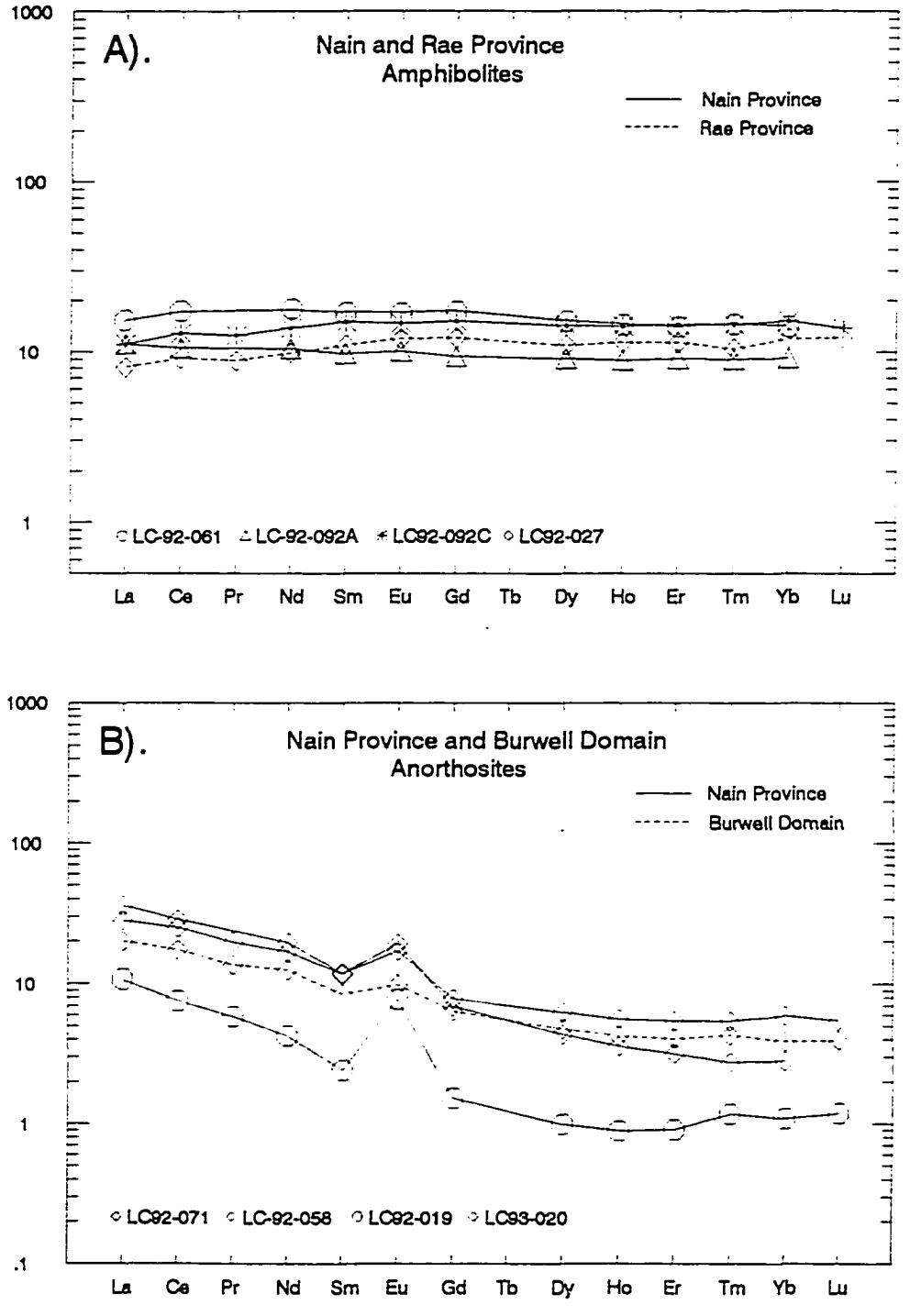


Figure 11. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). **(A)** - Amphibolites from the Nain and Rae provinces. **(B)** - Anorthosites from the Nain Province and Burwell Domain.

2.5.1.2 Rae Province

On CIPW normative diagrams, the Rae Province tonalitic rocks plot in the "trondhjemite-tonalite-granodiorite" field, and granitic rocks plot in the fields of "quartz monzonite" and "granite" (Fig. 4). The two groups differ in most aspects of their major and trace element geochemistry and will thus be addressed separately. Amphibolites from the Rae Province are tholeiitic and geochemically indistinguishable from amphibolites in the Nain Province.

Tonalitic Rocks

Al_2O_3 values for tonalitic rocks (i.e., tonalites, trondhjemites and granodiorites) are generally greater than 15 wt.% (Fig. 5A). They have low K_2O (1.4-2.1 wt.%), Ba (210-720 ppm) and Sr (120-540 ppm; Figs. 6 and 7). Na_2O/K_2O is low (1.9-3.1), Rb/Sr is variable (0.14-0.58), but relatively high compared with Nain Province tonalites and K/Rb is low (160-250) (Fig. 6).

Chondrite-normalized REE patterns are very similar among the tonalitic rocks (Fig. 12A), with strong REE fractionation (La_N = 65-100, Yb_N = 0.6-2.8, La_N/Yb_N = 48-128), and negligible Eu anomalies (Eu/Eu* = 0.8-1.4). On the MORB-normalized spider diagram (Fig. 13A), samples are enriched in Rb, Ba and Pb, similar to Nain Province tonalitic rocks. Rae Province tonalitic rocks are more strongly enriched in K relative to the Nain rocks, and they show a large positive Th spike, in contrast to the negative Th spike observed in most Nain Province tonalites. U and Sr are depleted relative to adjacent elements, but with values greater than MORB. Nb and Zr are slightly enriched (relative to adjacent elements) at ca. 1 to 3 times MORB. Ti, Y and Yb are strongly depleted relative to MORB.

Granite-quartz monzonite

Rae Province granitic rocks are characterized by 14-15.5 wt.% Al_2O_3 (i.e., slightly lower than for the tonalites at a given SiO_2; Fig. 5A), and high K_20 and Ba (4.14-5.78 wt.% and 1700-

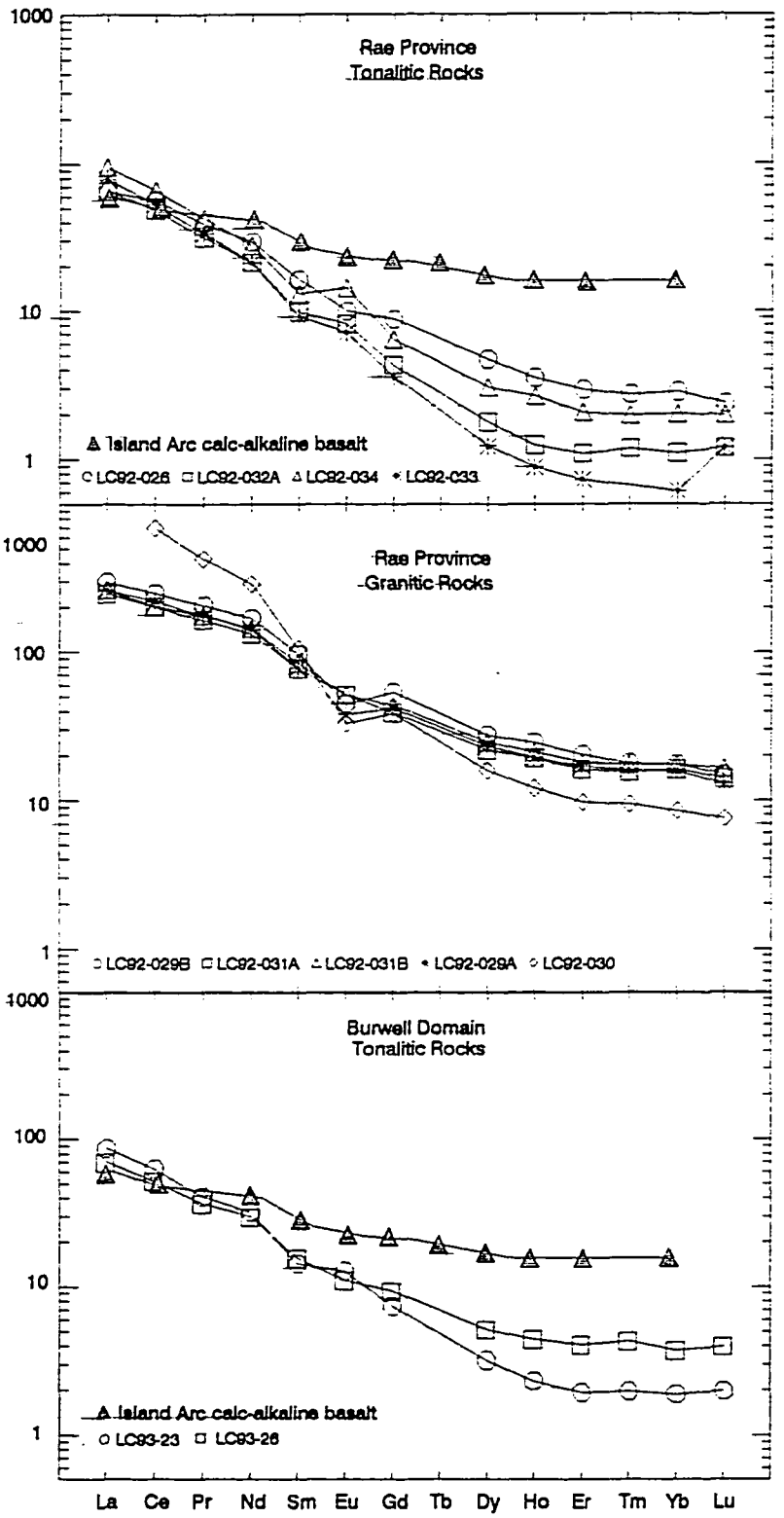


Figure 12. Chondrite normalized REE patterns for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). **(A)** - Tonalitic gneisses from the Rae Province. **(B)** - Granitic rocks from the Rae Province. **(C)** - Tonalitic gneisses from the Burwell Domain. Data for island arc calc-alkaline basalts from Wilson (1989).

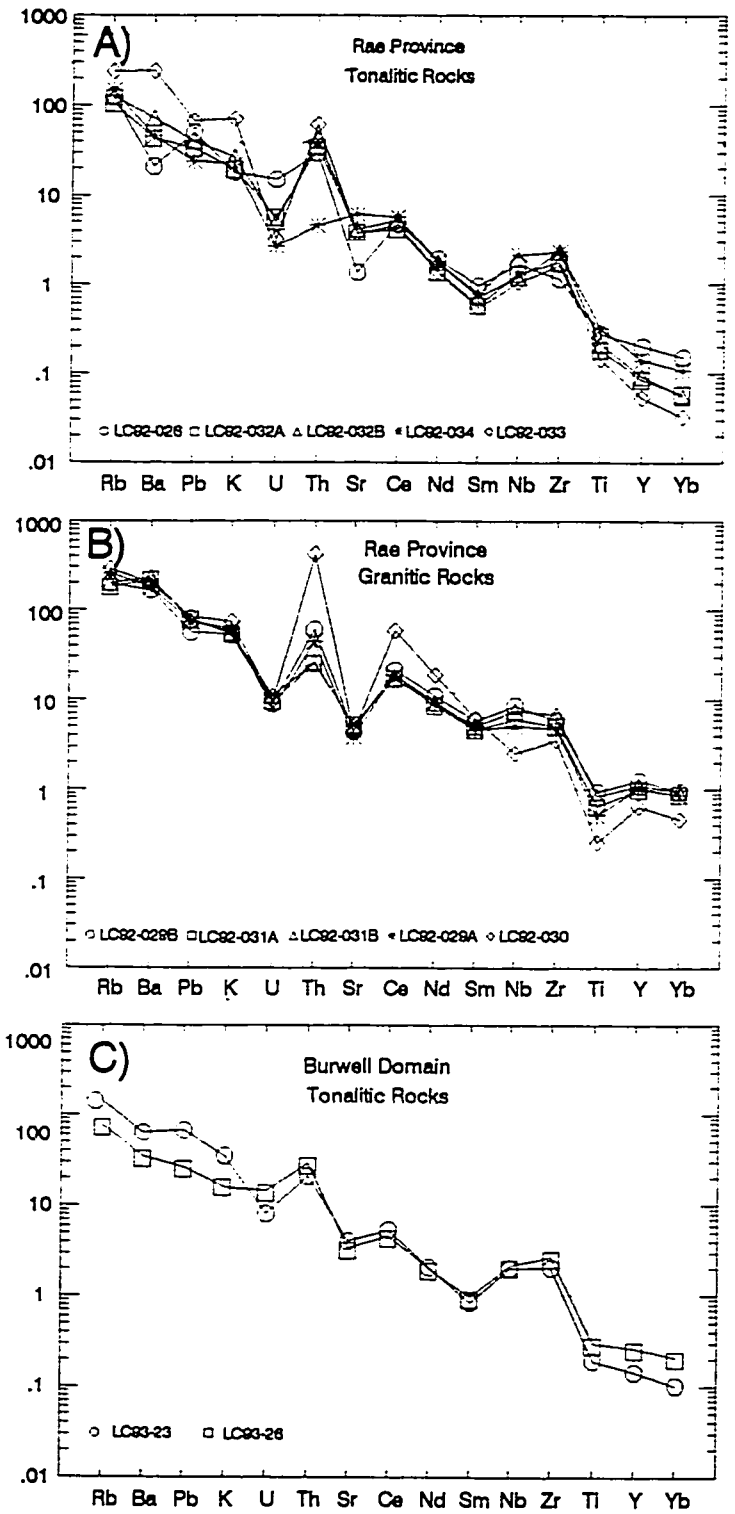


Figure 13. MORB normalized trace element data (Spider diagrams) for Archean rocks from the Torngat Orogen, Northern Labrador. Normalizing values and sequence of elements from McCulloch and Gamble (1990). **(A)** - Tonalitic gneisses from the Rae Province. **(B)** - Granitic rocks from the Rae Province. **(C)** - Tonalitic rocks from the Burwell Domain.

2300 ppm, respectively; Figs. 6 and 7). Rb/Sr and K/Rb are high (0.14-0.58 and 290-350, respectively), and Na_2O/K_2O is low relative to the tonalites (0.5-0.82; Fig. 6). The granites have higher overall REE abundances than the tonalites (Fig. 12B), with less severe REE fractionation (La_N = ca. 300, Yb_N = 10-17.5, La_N/Yb_N = 16-20), negative Eu anomalies (Eu/Eu* = 0.46-0.88) and relatively high concentrations of HREE (ca. 10-20 times chondrite; Fig. 4A). On the MORB-normalized spider diagram (Fig. 13B), the granites are strongly enriched in most elements (Rb, Ba, Pb, K and Th are ca. 80-300 times MORB). U and Sr are depleted relative to adjacent elements at 3-10 times MORB, and Ti, Y and Yb are slightly below MORB values.

Amphibolite

Rae Province amphibolite (only one sample, LC92-27), is similar to Nain Province amphibolites in major and trace element characteristics (Figs. 5-7, 11A and 10B). In the Rae Province, amphibolites are the oldest component present, and they locally contain rafts and layers of metasedimentary material and are always intruded by the tonalites. Sample LC92-27 is tholeiitic (Fig. 4A) and shows a flat REE pattern at concentrations ca. 10 times chondrite (Fig. 11A). Differences in chemical composition relative to Nain Province amphibolites include slightly higher SiO_2 and K_2O (Fig. 6).

2.5.1.3 Archean Gneisses In The Burwell Domain

On CIPW normative diagrams (Figs. 4B and C), the samples plot as tonalite (LC92-26), granodiorite (LC92-23) and anorthosite (LC92-20, a concordant layer in granodiorite LC92-23).

The range in major and trace element compositions is comparable to both Nain and Rae Province tonalites (Figs. 5, 6 and 7). REE patterns are also similar, (Fig. 12C) and show LREE enrichment (La_N = ca. 70-90), relative HREE depletion (Yb_N = ca. 2-4), and moderate REE fractionation (La_N/Yb_N = ca. 18-45). Eu anomalies are negligible (Eu/Eu* = 0.89-1.18). On the

MORB-normalized spider diagram (Fig. 13C), Burwell tonalites are similar to both Nain and Rae Province equivalents in most elements (enrichments in Rb, Ba, Pb and K), whereas the postive Th spike is only seen among Rae Province tonalites. Nb and Zr range from 1-3 times MORB. Ti, Y and Yb are strongly depleted relative to MORB.

2.5.2 Isotope Geochemistry of the Archean Rocks

Neodymium, strontium and lead isotopic data for selected whole rock samples are summarized in Table 3 and Figs. 14 through 16. A depleted mantle growth curve by DePaolo (1981) is used as a reference for the Nd isotopic evolution (Fig. 14), and a Stacey-Kramers two stage evolution curve (Stacey and Kramers, 1975) is used for reference to the $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ data (Fig. 16A).

2.5.2.1 Nain Province

Tonalites, trondhjemites and granodiorites

Nain Province tonalites, trondhjemites and granodiorites have $\varepsilon Nd_{(0)}$ values between -35.95 and -40.93, with low $^{147}Sm/^{144}Nd$ (0.0670-0.1031; Fig. 14). Samples LC92-1, 2 and 45 (granodiorite, trondhjemite and tonalite, respectively) form a cluster in Fig. 14B, but LC92-86 has unusually low $^{147}Sm/^{144}Nd$ relative to the other samples.

Measured $^{87}Sr/^{86}Sr$ ratios for Nain Province tonalitic rocks are 0.702914-0.709453, with $^{87}Rb/^{86}Sr$ values of 0.0319-0.1578 (Fig. 15A). The unradiogenic $^{87}Sr/^{86}Sr$ is also a characteristic feature of Nain Province gneisses occurring in the Saglek block further to the south (Schoitte et al., 1990; Schoitte and Bridgwater, 1991).

On an $\varepsilon Nd_{(0)}$ vs. $^{87}Sr/^{86}Sr_{(0)}$ diagram (Fig. 15B), Nain Province tonalitic samples show a negative linear correlation, indicating that the two isotope systems have generally remained correlated, despite the likelihood of open system behavior since crystallization of the samples.

TABLE 3 - Isotopic analyses of Archean gneisses, Torngat Orogen, Northern Labrador

	Rb (ppm)	Sr (ppm)	87Rb/ 86Sr	87Sr/ 86Sr(M)	Sm (ppm)	Nd (ppm)	147Sm/ 144Nd	143Nd/ 144Nd(M)	143/144Nd (±2 sigma)	ENd (M)	T(DM) (Ga)	206Pb/ 204Pb	207Pb/ 204Pb	208Pb/ 204Pb
NAIN PROVINCE														
Tonalitic rocks														
LC92-001	20.6	427	0.1398	0.70945	1.66	10.6	0.0945	0.510540	14	-40.93	3.24	13.470	14.547	33.425
LC92-45		609		0.70326	3.70	21.7	0.1031	0.510769	5	-36.46	3.18	13.751	14.516	33.057
LC92-086	30.9	566	0.1578	0.70562	1.54	13.9	0.0670	0.510667	41	-38.45	2.48	13.594	14.505	33.191
LC92-88												13.623	14.545	33.248
Intermediate samples														
LC92-002	17.3	1563	0.0319	0.70291	15.39	97.5	0.0955	0.510795	34	-35.95	2.92	13.738	14.641	35.872
LC92-47B												13.511	14.459	33.078
Amphibolites														
LC92-061					3.26	10.6	0.1857	0.512445	10	-3.76	3.46	14.575	14.748	34.385
LC92-92A												14.256	14.675	34.304
Anorthosites														
LC92-19	43.9	642	0.1977	0.70362	0.33	2.1	0.0953	0.510986	7	-32.23	2.66	15.138	15.030	35.613
LC92-58					1.70	9.1	0.1130	0.511129	5	-29.44	2.93	14.002	14.657	34.091

Analytical details are given in Appendix 1.

TABLE 3 (continued) - Isotopic analyses of Archean gneisses, Torngat Orogen, Northern Labrador

	Rb (ppm)	Sr (ppm)	87Rb/ 86Sr	87Sr/ 86Sr(M)	Sm (ppm)	Nd (ppm)	147Sm/ 144Nd	143Nd/ 144Nd(M)	143/144Nd (±2 sigma)	ENd (M)	T(DM) (Ga)	206Pb/ 204Pb	207Pb/ 204Pb	208Pb/ 204Pb
RAE PROVINCE														
Tonalitic rocks														
LC92-26	53.2	115	1.3365	0.75898	2.82	17.2	0.0995	0.510770	8	-36.44	3.07	16.136	15.319	35.749
LC92-32a					1.32	9.1	0.0879	0.510550	29	-40.73	3.05	14.732	14.946	35.698
LC92-33												14.555	14.944	35.611
LC92-34												14.538	14.802	35.749
Granitic rocks														
LC92-29A												14.293	14.846	35.028
LC92-30												14.616	14.896	41.413
LC92-31												14.148	14.771	35.462
Amphibolites														
LC92-27		75		0.71227	1.70	4.7	0.2171	0.513078	18	8.58	2.00	16.131	15.360	35.630
LC92-28												16.331	15.299	44.779
BURWELL DOMAIN														
Tonalitic rocks														
LC93-23	65.1	335	0.5617	0.72557	2.07	13.6	0.0917	0.510877	10	-34.35	2.72	15.238	15.013	34.993
LC93-26	32.1	251	0.3697	0.71706	2.26	12.9	0.1055	0.511070	9	-30.59	2.80			
Anorthosite														
LC93-20												15.911	15.096	35.565

Analytical details are given in Appendix 1.

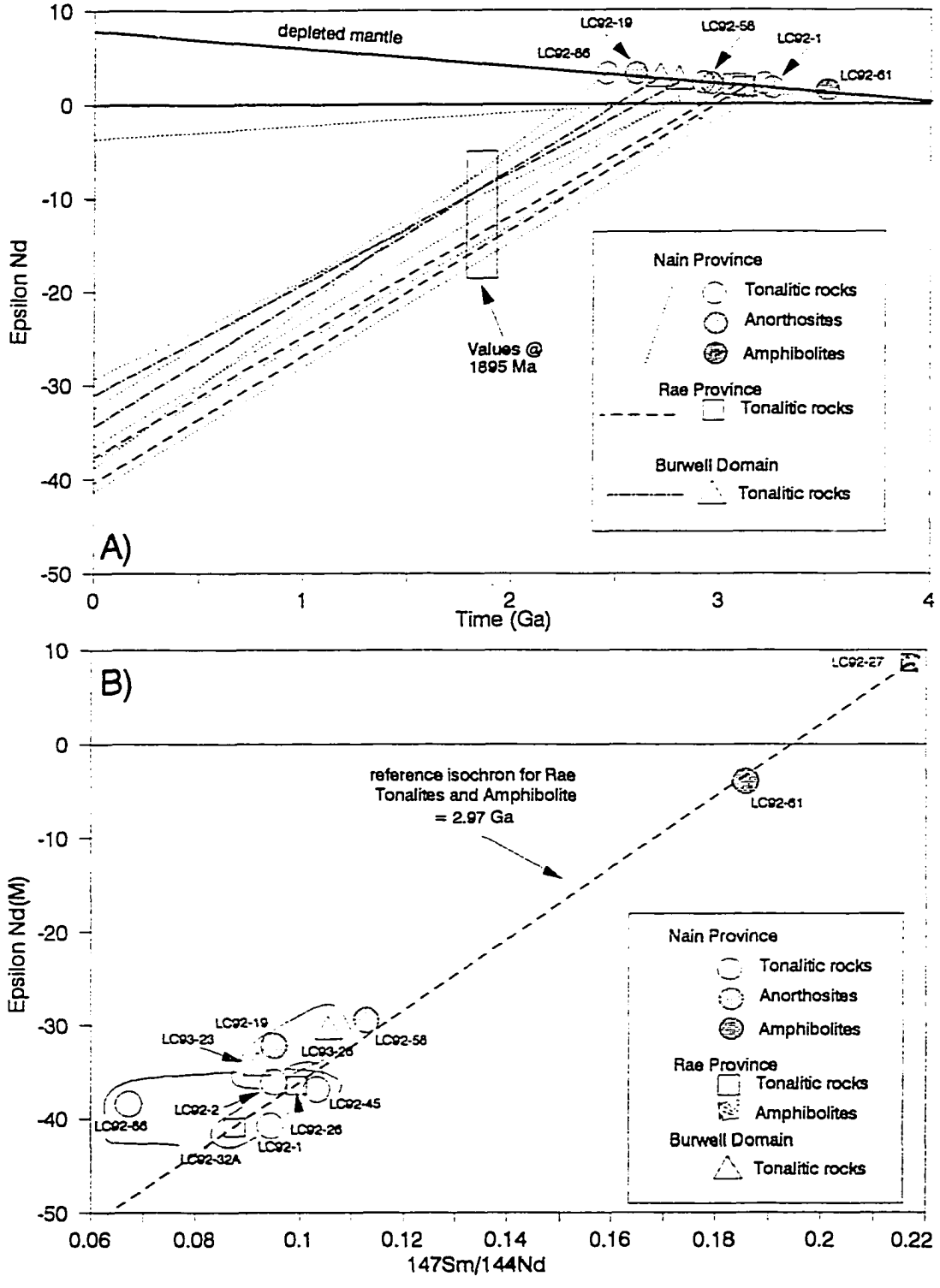


Figure 14. Sm/Nd isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. **(A)** - εNd vs. age diagram showing Nd isotopic evolution lines for a range of lithologies. Depleted mantle curve from DePaolo (1981). **(B)** - εNd (M) vs. $^{147}Sm/^{144}Nd$ for a range of lithologies. Reference isochron of 2.97 Ga is based on samples from the Rae Province (see also text).

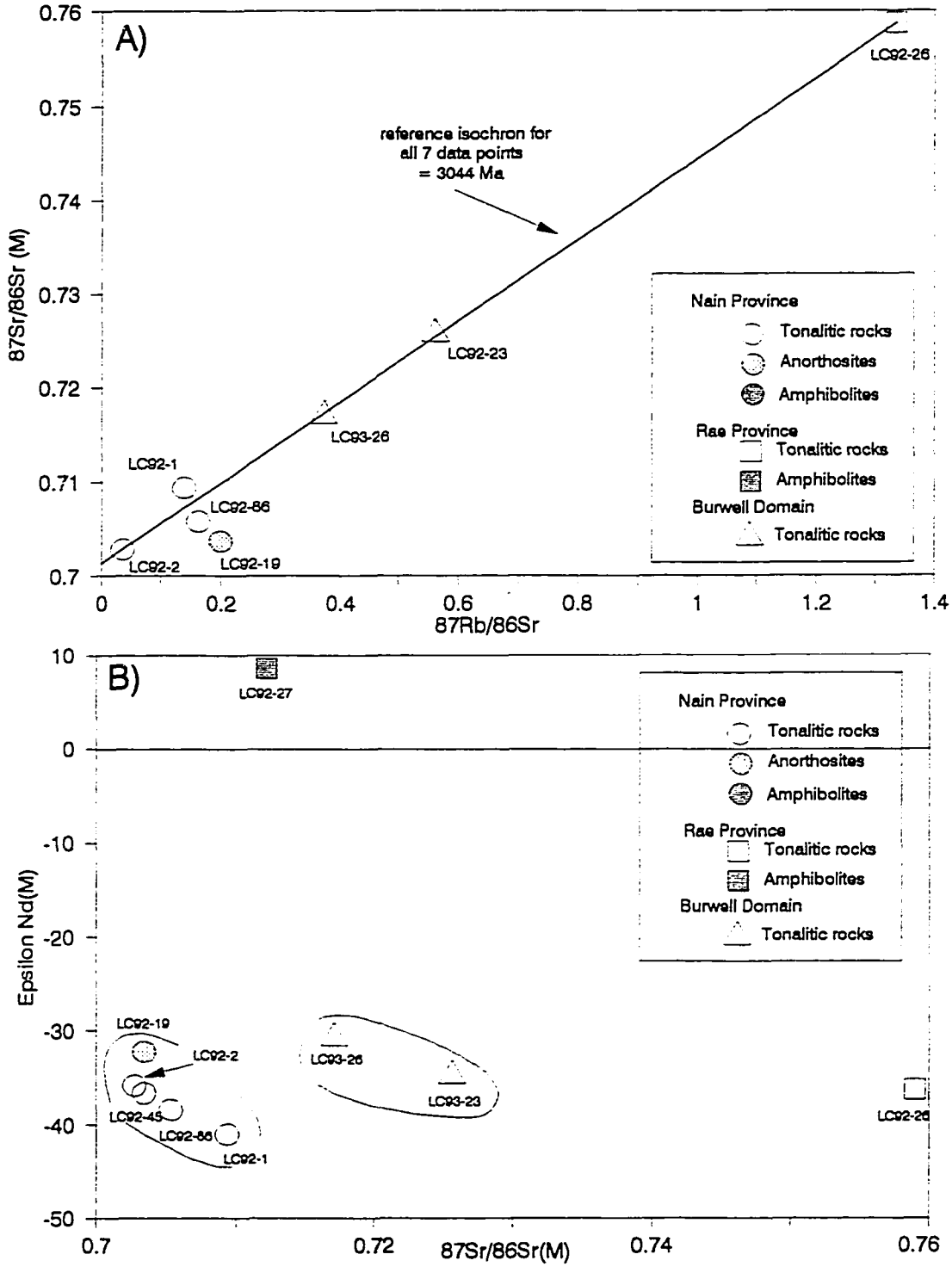


Figure 15. Rb/Sr and Sm/Nd isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. (A) - Rb/Sr isochron diagram for all samples analyzed. Reference isochron of 3044 Ma is based on all data points. (B) - $^{87}Sr/^{86}Sr(M)$ vs. εNd (M) diagram for all samples analyzed.

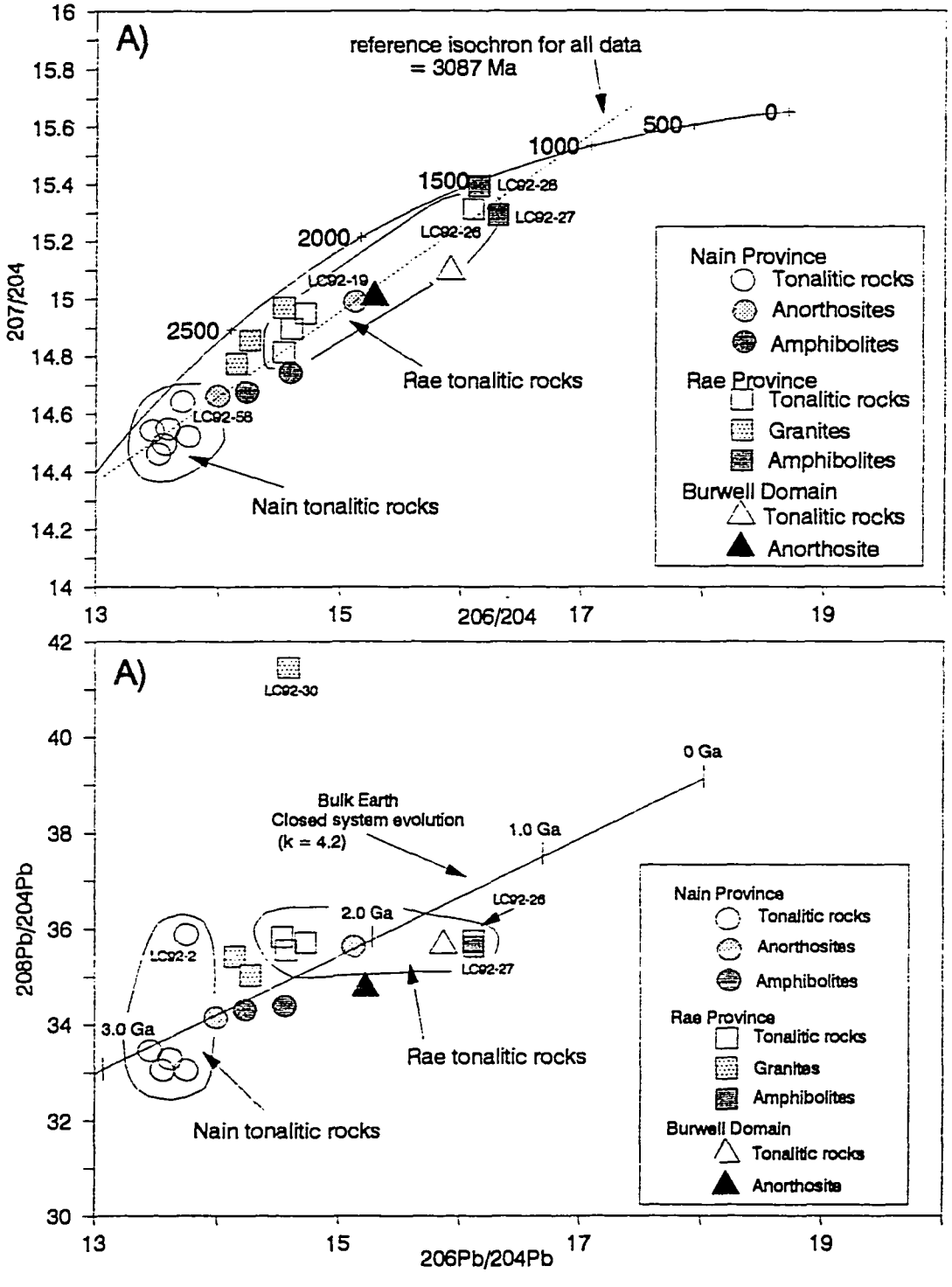


Figure 16. Pb isotope data for Archean rocks from the Torngat Orogen, Northern Labrador. **(A)** - $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ isotope correlation diagram. Reference isochron of 3087 Ma is based on all data. Stacey-Kramers two stage evolution curve (Stacey & Kramers, 1975) is shown for reference. **(B)** - $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ isotope correlation diagram. The "Bulk earth closed system evolution" curve (k = 4.2) is from Gariépy and Dupré (1991) and is shown for reference.

The Nain Province tonalitic rocks have relatively unradiogenic $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ ratios (see later comparisons), showing only little variation ($^{206}Pb/^{204}Pb$ = 13.470-13.751; $^{207}Pb/^{204}Pb$ = 14.459-14.641; Fig. 16A). $^{208}Pb/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ also form a tight cluster with unradiogenic values (33.078-33.425), with the exception of LC92-2, which has unusually high $^{208}Pb/^{204}Pb$ relative to other tonalitic samples from the Nain Province (35.872) (Fig. 16B).

Amphibolites and anorthosites

Amphibolite LC92-61 yields $\varepsilon Nd_{(0)}$ of -3.76 and $^{147}Sm/^{144}Nd$ of 0.1857 (Figs. 14A and B). Pb isotope data for LC92-61 and LC92-92A have more radiogenic $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ values than the tonalites (Figs. 16A and B; Table 2). Anorthosites LC92-58 and LC92-19 were collected from different segments along the same linear belt (Fig. 1). They have $\varepsilon Nd_{(0)}$ of -29.44 and -32.23, and $^{147}Sm/^{144}Nd$ of 0.11299 and 0.0953, respectively (Figs. 14A and B). Pb ratios for the two samples are dissimilar, but they plot along the 3087 Ma secondary isochron which is defined by many of the Archean samples from the northern TO in this study (Fig. 16A, Table 2). LC92-19 has unradiogenic $^{87}Sr/^{86}Sr$, which is similar to that of the tonalites ($^{87}Sr/^{86}Sr$ =0.70362), but at a higher $^{87}Rb/^{86}Sr$ (0.1977; Fig. 15A).

2.5.2.2 Rae Province

Tonalites, Trondhjemites, Granodiorites

Two tonalites from the Rae Province (LC92-26 and LC92-32A) yield $\varepsilon Nd_{(0)}$ = -36.44 and -40.73, with $^{147}Sm/^{144}Nd$ = 0.0995 and 0.0879, respectively (Figs 14A and B). $^{87}Sr/^{86}Sr$ for LC92-26 is strongly radiogenic relative to those of Nain Province tonalites ($^{87}Sr/^{86}Sr_{(0)}$ = 0.758979 and $^{87}Rb/^{86}Sr$ = 1.3365; Fig. 15). $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ ratios for Rae Province tonalites are also more radiogenic than Nain Province equivalents. $^{208}Pb/^{204}Pb$ values are constant among all Rae Province samples over a large range in $^{206}Pb/^{204}Pb$ (Fig. 16B ; $^{208}Pb/^{204}Pb$ = 35.611-35.749).

Granite- Quartz Monzonite

There is currently no Nd or Sr isotope data for the granitic rocks from the Rae Province. Pb ratios are almost indistinguishable from the values for the tonalities they intrude, but the granites generally have slightly lower $^{206}Pb/^{204}Pb$ for a given $^{207}Pb/^{204}Pb$ or $^{208}Pb/^{204}Pb$ (Figs. 16A and B). Sample LC92-30 has an anomalously high $^{208}Pb/^{204}Pb$ relative to the other granites (Fig. 16B).

Amphibolite

Amphibolite LC92-27 has $\varepsilon Nd_{(M)} = 8.58$, with $^{147}Sm/^{144}Nd$ of 0.2171 (Figs. 14A and B). The sample has strongly radiogenic Pb ($^{206}Pb/^{204}Pb = 16.131$, $^{207}Pb/^{204}Pb = 15.360$ and $^{208}Pb/^{204}Pb = 35.630$), with values similar to the tonalite (LC92-26), which intrudes it (Figs. 16A and B).

2.5.2.3 Archean Gneisses in the Burwell Domain

The two Archean tonalites from the Burwell Domain (LC93-23 and LC93-26), yield $\varepsilon Nd_{(M)}$ values slightly higher than those of Nain or Rae Province tonalitic rocks (Figs. 14A and B). $^{87}Sr/^{86}Sr_{(M)}$ and $^{87}Rb/^{87}Sr$ values are higher than Nain Province tonalites and lower than Rae equivalents (Fig. 15A). Pb ratios for a tonalite, and for a concordant anorthosite sheet within it are much more radiogenic than Nain Province samples, and are generally more radiogenic than Rae Province samples (Figs. 16A and B).

2.6 DISCUSSION

2.6.1 Tonalitic gneisses in northern Labrador: Comparisons with high-Al TTDs

Tonalitic gneisses from the Nain and Rae provinces and the Burwell Domain are all characterized by high Al_2O_3 (> 15 wt.%), high Sr (> 300 ppm), low Rb/Sr (< 0.15), LREE enrichment, and low K/Rb (< 550), Nb (< 11 ppm) and Y (< 11ppm). Felsic samples (SiO_2 > 65 wt.%), further show HREE depletion (relative to LREE) and, in the case of Nain Province

tonalites, a positive Eu anomaly. These characteristics differ significantly from average Phanerozoic calc-alkaline rocks which typically have higher K_2O, less depleted HREE and a negative Eu anomaly; compositions that are best explained by differentiation at shallower crustal levels (< 40 km), where Ca-plagioclase, rather than garnet, is a dominant residual phase (Taylor and McLennan, 1985).

The geochemical characteristics summarized above are identical to those of Archean "high-Al TTD" (tonalites, trondhjemites and dacites - the extrusive equivalent of granodiorites), as originally defined by Barker and Arth (1976) and further discussed by e.g., Glikson (1979); Martin (1986); Drummond and Defant (1990) and Winther and Newton (1990). Geochemical modeling of the high-Al TTD rocks (Arth, 1976; Martin, 1986) suggests that they formed by partial melting of either basaltic amphibolite or eclogite with garnet and amphibole +/- clinopyroxene in the residue. This implies that partial melting occurred at depths below the plagioclase stability field, and that the basalt source was transformed to eclogite or to garnet-amphibolite prior to partial melting (e.g., Martin, 1986; Winther and Newton, 1989; Drummond and Defant, 1990). Such a transformation requires a high P-T environment which could be achieved either during subduction of basaltic oceanic crust or by basaltic underplating at the base of the crust, followed by crustal thickening (Kay and Kay, 1991).

The high Al_2O_3 concentrations observed in typical Archean high-Al TTDs (ca. 15 wt. % at 70 % SiO_2), have been suggested by Drummond and Defant (1990) to be caused by separation of the melt from a residue enriched in sub-aluminous hornblende (these are shown to co-exist at pressures of 5-10 Kb; Ellis and Thompson, 1986). The HREE and Y depletion (and resulting high $La_{(N)}/Yb_{(N)}$) observed in high-Al TTDs also support separation of the melt from a residue enriched in garnet and/or hornblende, as these minerals have high Kds for HREEs and Y, and therefore efficiently remove these elements from the melt (Drummond and Defant, 1990). Similarly, the high Sr concentrations observed in high-Al TTDs can be explained by low Sr Kds of residual hornblende, garnet and clinopyroxene. The low K/Rb values are thought to be a function of

hornblende extraction because of its greater affinity for K than for Rb (Glikson, 1976) and the low concentrations of Nb and other HFSEs among high-Al TTDs are argued to result largely from the strong affinity of residual hornblende for these elements (Drummond and Defant, 1990).

The model by Martin (1993) summarizes the above discussion on the origin of high-Al TTD magmas: (a) mantle derived, garnet-amphibolitic oceanic crust undergoes partial melting, possibly during subduction, (b) the resulting tonaltic partial melt undergoes fractional crystallization (primarily of hornblende) to produce a residue of layered mafic rocks and intermediate to felsic melts (the high-Al TTD suites), which (c) intrudes at higher levels.

2.6.2 Significance of Eu-anomalies

Positive Eu anomalies are not typical of all Archean high-Al TTD rocks (Arth and Barker, 1976). Within the Torngat region positive Eu anomalies are only observed in some of the Nain Province tonalitic gneisses with $SiO_2 > 65$ wt.% (Fig. 7E, 8A-C) and do not correlate with higher Sr or CaO (Figs. 5D, 6B and 11), suggesting that the Eu anomalies were not entirely controlled by plagioclase accumulation. Several of the samples with large positive Eu anomalies also show extreme REE fractionation, suggesting that for these samples, the positive Eu anomalies were enhanced by separation of the melt from residual phases such as garnet and hornblende, which are enriched in HREEs, have negative Eu anomalies and therefore contribute to the production of melts with HREE depletion and a positive Eu anomaly (Arth and Barker, 1976; Henderson, 1984; Drummond and Defant, 1990).

2.6.3 Geochemical Effects Of Granulite Facies Metamorphism

The Archean rocks analyzed in this study have experienced extended and complex metamorphic histories. While Early Proterozoic high grade metamorphism associated with Torngat orogenesis is reasonably well-constrained by recent field, structural, metamorphic, and geochronological studies, the grade and extent of Archean metamorphism is more difficult to

assess. The Archean basement gneisses in the Nain and Rae Provinces and Burwell domains are all known to have experienced at least one phase of upper amphibolite to granulite facies prograde metamorphism prior to Early Proterozoic orogenesis (D. Scott, pers. comm, 1997; Connelly and Mengel, 1996, Schiotte et al., 1990). Subsequently, the Archean basement gneisses were metamorphosed during collision along the Torngat Orogen, which was followed by variable retrograde metamorphism associated with unroofing of the rocks and movement along the KSZ.

The geochemical effects of metamorphism can best be evaluated by comparing metamorphosed rocks of known protolith composition with their unmetamorphosed equivalents. The Nain Province amphibolites provide the most likely protolith compositions in this study because their geochemical and isotopic characteristics suggest that they represent metamorphosed MORB (see above). Field relationships indicate that the Nain Province amphibolites were intercalated with Nain Province basement gneisses during the Late Archean or Early Proterozoic. Therefore their use as monitors of element mobilities associated with metamorphism is limited to the Early Proterozoic metamorphic events.

Relative to MORB, the amphibolites are enriched in incompatible elements Rb, Ba, Pb, (+/- K), but notably not in LREEs, Sr, U or Th (which would liken them to island arc basalts), and not in Nb through Yb (Fig. 10B). On the basis of the similarity of most element concentrations to those of MORB, the amphibolites are interpreted as MORBs which were subsequently slightly enriched in LILEs by infiltrating fluids (McCulloch and Gamble, 1990; Touret, 1996). The amphibolites occur in the Komaktorvik shear zone (Fig. 2), which is known to have experienced stronger retrograde metamorphism (associated with late-stage deformation under amphibolite facies conditions: Van Kranendonk et al., 1994; Rivers et al., 1996), than the neighboring gneisses. We therefore consider it likely that the LILE-enrichment observed in the amphibolites reflect this setting, and may be more pronounced than elsewhere in the northern TO.

An additional method used to evaluate element mobilty associated with high grade metamorphism is Rb/K ratios. Rb/K ratios are a good indicator of element depletion because Rb

commonly responds more sensitively to granulite facies metamorphism than K (e.g., Rudnick et al., 1985; Taylor and McLennan, 1985; Touret, 1996). This is particularly true for low potassium rocks (< 1 wt.%), in which biotite (the dominant K- and, especially, Rb-bearing phase), easily breaks down during high grade metamorphism, thus causing preferential Rb loss. In more potassic rocks, K-feldspar is the dominant K- and Rb-bearing phase, but this mineral has lower Rb/K than biotite, it is stable under granulite facies conditions and it is less sensitive to Rb loss. Thus, during granulite facies metamorphism, Rb/K ratios decrease more in low-K rocks (breakdown of Rb-rich biotite) than in high-K, K-feldspar bearing rocks.

The Archean gneisses in this study are plotted on Figure 17 in terms of Rb (ppm) vs. K (wt. %). The samples do not show the preferential depletion of Rb relative to K, as is common among rocks that have been metamorphosed under granulite facies conditions (Rudnick et al., 1985). In contrast, Figure 17 shows that most samples plot within the field of unmetamorphosed igneous rocks (samples showing Rb contents that are slightly high relative the main trend of unmetamorphosed igneous rocks are anorthosites). The unradiogenic $^{87}Sr/^{86}Sr$ values and low Rb/Sr ratios seen among Nain Province gneisses further suggest that secondary enrichment of Rb was insignificant in these rocks (Fig. 15A). Thus, on the basis of the data shown on Figures 15A and 17, LILE-depletion associated with the ca. 2.7 Ga granulite facies metamorphism within the Nain Province did not cause a substantial shift in Rb/K, and secondary enrichment of LILE element did not significantly alter the chemical compositions or the Sr isotopic character of these rocks.

2.6.4 The composite nature of the Nain Province

Nain Province tonalitic rocks have strongly negative $\varepsilon Nd_{(t)}$ values (Fig. 14). U-Pb zircon analyses for a few rocks from the area indicate crystallization ages between 2834 and 2650 Ma (Scott, 1995b). Although the limited age data set does not necessarily reflect the entire spectrum of Archean crustal ages from the northern segment of the Nain Province, it does establish the

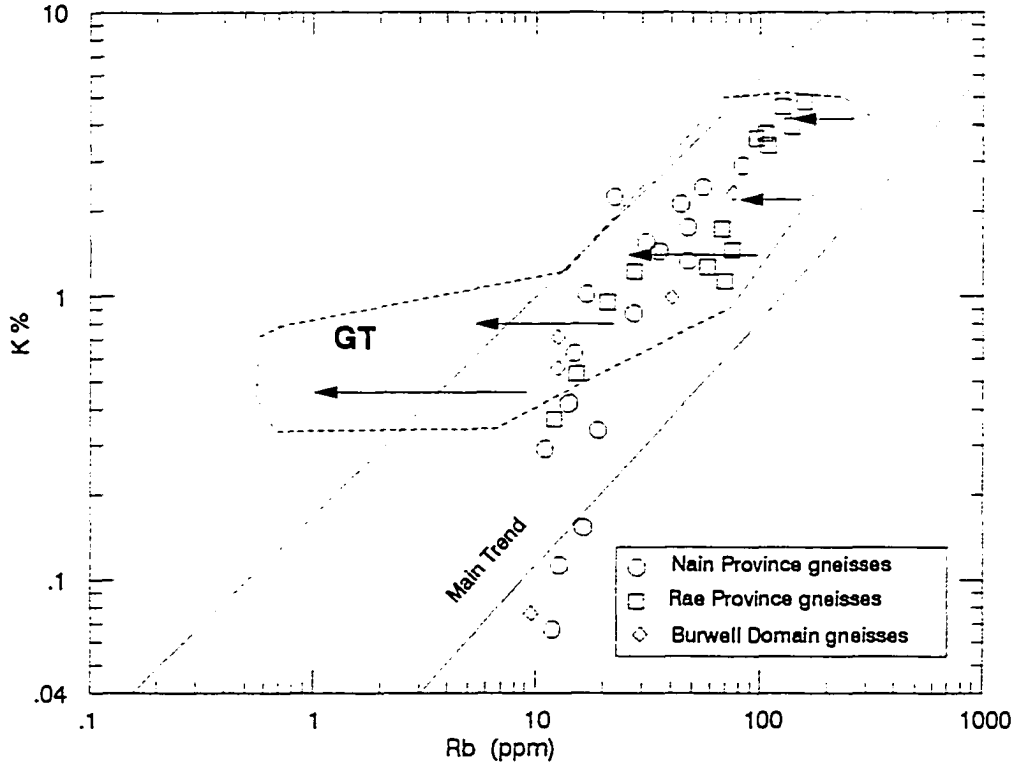


Figure 17. Plot of Rb (ppm) versus K (wt.%) for Archean tonalitic, trondhjemitic and granodioritic gneisses in the northern Torngat Orogen. The diagram also shows the "Granulite Trend" (GT) from Rudnick et al. (1985) and the "Main Trend" of continental igneous rocks (granites through basalts) from Shaw (1968). Horizontal arrows represent the amount and direction of Rb depletion caused by granulite facies metamorphism (from Rudnick et al., 1985).

existence of Mid- to Late Archean ages, and the low $\varepsilon Nd_{(0)}$ values reported here are thus interpreted to reflect the time-integrated isotopic evolution of the Sm-Nd system.

The tonalitic gneisses have low $^{87}Rb/^{86}Sr$ and $^{238}U/^{204}Pb$ ratios coupled with unradiogenic $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$ ratios (Figs. 15 and 16), which reflect long-lived U and Rb depletions. Similar U and Rb depletion associated with ca. 2.7-2.8 Ga granulite facies metamorphism is documented ca. 160 km to the south in the Saglek block, as well as elsewhere in the North Atlantic Archean craton (e.g., Bridgwater et al., 1973; Schiotte, 1988; Schiotte et al., 1990; Bridgwater and Schiotte, 1991). In the current study area, Scott (1995b) and Connelly and Mengel (1995) have identified 2761-2770 Ma metamorphic zircons, indicating that granulite facies metamorphism occurred synchronously with that in the Saglek block. It is therefore reasonable to conclude that the U and Rb depletions observed in the study area resulted from this high grade metamorphic event.

The oldest crust identified in the Four Peaks Domain is from Seven Islands Bay (Fig. 1; ca. 3.75 Ga on the basis of Pb-Pb data, Bridgwater & Wardle, 1992) and represents an age similar to the Uivak gneisses in Saglek fiord (up to 3.7-3.9 Ga; e.g., Bridgwater and Schiotte, 1991; Schiotte et al., 1991). While this occurrence is of very limited extent, it does attest to the composite nature of the Nain crust.

The Nain Province tonalitic gneisses which dominate the study area north of Eclipse Harbour are much more similar to the Mid- to Late Archean Parkavik gneisses from the Okak area (Fig. 1; Schiotte et al., 1990), based on comparable U-Pb zircon ages and U-Rb depletions associated with 2.7-2.8 Ga granulite facies metamorphism. In contrast, other Mid Archean gneisses from the Saglek and Hopedale blocks (the Lister and Maggo gneisses, respectively), do not record the distinctive U and Rb depletion and are thus considered to represent shallower crustal segments which did not experience granulite facies metamorphism (Bridgwater and Schiotte, 1991, Schiotte et al., 1991, Bridgwater and Wardle, 1992).

2.6.4.1 Origin of the anorthosites

Crystallization ages for the anorthosites are presently not known directly, however, based on field criteria discussed above, they are likely to have been juxtaposed with more Nain Province gneisses during Latest Archean or Earliest Proterozoic. With the relatively high $^{147}Sm/^{144}Nd$ ratios of the anorthosites, their low $\varepsilon Nd_{(0)}$ values require either a very long time-integrated isotopic evolution, or a considerable degree of contamination by older crust (Fig. 14). Due to the lack of age determinations, it is currently not possible to constrain the degree of crustal contamination experienced by the anorthosites and therefore it is not possible to evaluate the environment into which they were emplaced. It has recently been proposed (e.g., Ashwal, 1993) that similar anorthosite complexes formed within oceanic crustal environments, based on their spatial association with voluminous basaltic lavas, for example in ancient anologs to a mid ocean ridge or back arc basin (Ashwal, 1993). A corollary of the above is that the present location of anorthosites (and associated amphibolites) in continental crust marks the presence of a former ocean.

2.6.5 Rae Province - comparison with the Nain Province

Rae Province tonalites are geochemically similar to those in Nain Province, but they have higher Rb and Th concentrations and higher Rb/Sr (Figs. 6D and E, 13A).

The $\varepsilon Nd_{(0)}$ and Sm/Nd values are indistinguishable from those in equivalent Nain Province rocks (Fig. 14). Preliminary U/Pb zircon ages of ca. 3.0 and 2.7 Ga from Rae Province gneisses (D. Scott, pers. comm., 1997), coupled with low $\varepsilon Nd_{(0)}$ values, indicate that Sm-Nd evolved in a relatively closed system and was not significantly disturbed during Proterozoic orogenesis. Rae Province samples produce a linear array along a 2.97 Ga reference isochron (Fig. 14B), indicating that the Sm/Nd ratios were not disturbed since that time. This implies that crustal reworking associated with Early Proterozoic collision did not significantly disturb the Nd isotopic systematics in this area.

The higher Rb/Sr and distinctively more radiogenic $^{87}Sr/^{86}Sr$ ratios (relative to Nain Province tonalites; Figs. 6 and 15A), suggests that rocks comprising this part of the Rae province did not experience severe LILE depletion associated with ca. 2.7 Ga granulite facies metamorphism, as did the neighboring Nain province.

Similarly, Pb isotopic ratios are more radiogenic than those from the Nain Province (Fig. 16A,B). The relatively constant $^{208}Pb/^{204}Pb$ ratios over a large range in $^{206}Pb/^{204}Pb$ (Fig. 16B), suggests a relatively recent disturbance to Th/Pb ratios in Rae Province gneisses. This is most likely related to the distinctive Th spike (Fig. 13) that characterizes all Rae Province samples in this study, suggesting that Th enrichment occurred in response to metamorphic events associated with Torngat orogenesis.

2.6.5.1 Age and origin of granitic rocks in the Rae Province

The granitic rocks that intrude Rae Province tonalitic gneisses are geochemically unique to the study area and clearly do not resemble high-Al TTDs. Relative to the tonalites they intrude, they have high concentrations of K_2O, Rb and HFSEs. Overall REE concentrations are also relatively high, the REEs are not strongly fractionated, and they have negative Eu anomalies. (Figs. 5-7, 12B and 13B). On the basis of these characteristics, the granites are considered to represent crustal differentiates which formed by anatexis of underlying Rae Province gneisses at shallow crustal levels (< 40 km), where An-rich plagioclase, rather than garnet, is a dominant residual phase (e.g., Taylor and McLennan, 1985). However this suggestion requires further substantiation by Nd isotopic analyses of the granites.

The granites have Pb isotopic ratios similar to those of the tonalites they intrude (Figs. 16A and B). This further supports the suggestion that localized Pb isotopic homogenization is associated with migmatization of Rae crust. There are currently no direct age constraints on the granites in this study, however monazite grains from adjacent tonalitic gneisses (LC92-26 and -32A), yield metamorphic U-Pb ages of 1835-1839 Ga (D. Scott, pers. comm., 1997), which overlap

with ages for transcurrent shearing along the ASZ following Nain-Rae collision (1845-1822 Ma; Bertrand et al., 1993). In addition, Bardoux et al. (1996) identified 1853 Ma migmatites with a similar field description, occurring ca. 100 km to the southwest (Fig. 3). It is thus reasonable to assume that localized crustal reworking and migmatization within the Rae Province accompanied ca. 1839-1835 Ma deformation associated with Torngat orogenesis.

2.6.5.2 Origin of Amphibolites in the Rae Province

Trace element concentrations for amphibolite LC92-27 are indistinguishable from those of Nain Province amphibolites and are similar to N-MORB (Fig. 10B). The sample has $^{147}Sm/^{144}Nd$ and $\varepsilon Nd_{(0)}$ values greater than chondrite, suggesting either a very depleted mantle source, or that the amphibolite represents the residue from which a partial melt was extracted, however the lack of a significant LREE depletion renders the latter interpretation less likely. $^{87}Sr/^{86}Sr$ is unradiogenic (Fig. 15B), whereas Pb isotopic ratios are strongly radiogenic with values identical to the tonalite that intrudes it (LC92-26; Fig. 16A), suggesting Pb isotopic homogenization, at least on the outcrop scale.

2.6.6 Burwell Domain - links to the Rae and Nain Province?

Burwell Domain tonalite gneisses resemble high-Al TTD gneisses discussed above. They are enriched in Th and do not have a Eu anomaly (Figs 12C and 13C), thus they are geochemically more similar to Rae Province tonalites than to those of the Nain Province.

$\varepsilon Nd_{(0)}$ values are just slightly higher than those of Nain and Rae Province equivalents, with identical $^{147}Sm/^{144}Nd$ values (Fig. 14). Preliminary U-Pb data from metamorphic zircons and monazites from these rocks indicate metamorphism at 2772 Ma (Connelly and Mengel, 1995), so their crystallization ages approximate those of compositionally similar tonalites occurring in the Nain and Rae provinces.

$^{87}Sr/^{86}Sr$ values are intermediate between the less radiogenic Nain gneisses and the more radiogenic Rae gneisses. The $^{87}Sr/^{86}Sr$ data show a positive correlation with $^{87}Rb/^{86}Sr$ and when combined with data for Rae and Nain tonalites, they plot along a 3044 Ma reference isochron. The significance of this isochron age is not clear, but it suggests that the coupling between $^{87}Sr/^{86}Sr$ and $^{87}Rb/^{86}Sr$ is a long lived feature of the Archean Burwell basement as well as of the Nain and Rae Provinces (Figs. 15). Pb isotopic ratios are radiogenic relative to Nain Province tonalites and plot within the field of Rae Province tonalites.

2.6.7 Summary

The results of the discussion and considerations above can be summarized in the following points:

1. The Archean Nain, Rae and Burwell cratonic blocks were tectonically juxtaposed during Torngat orogenesis and now collectively, along with Proterozoic intrusives and metasediments, constitute the northern TO. These Archean gneisses are composed predominantly of high Al TTD grey gneisses (e.g., Barker and Arth, 1976), which most likely formed by partial melting (at depths below plagioclase stability) of basalt which was transformed into garnet amphibolite or eclogite.

2. Nain Province Archean gneisses from the Four Peaks Domain have characteristically unradiogenic $^{87}Sr/^{86}Sr$ and Pb isotopic ratios relative to analagous gneisses from the adjacent Rae Province and Burwell Domain. This feature reflects a long lived depletion in Rb and U, a characteristic that is also observed in Nain Province gneisses of the Saglek block, where it is attributed to granulite facies metamorphism accompanying ca. 2.8-2.7 Ga cratonic amalgamation (Bridgwater and Schiotte, 1991). The unradiogenic Sr and Pb isotope ratios, coupled with mid-Archean crystallization ages for rocks in the northern-most Nain Province (Scott, 1995b), suggest they are correlative to the Parkavik gneisses exposed in the Saglek block near Okak Fjord (Schiotte et al., 1990), as previously proposed by Bridgwater (pers. comm., 1997).

3. Rae Province tonalite gneisses exposed immediately south of the Burwell Domain yield ca. 3.0 - 2.7 Ga zircon ages (D. Scott, pers. comm., 1997) and have εNd values identical to the compositionally similar Nain Province tonalites. Nd isotopic data alone do not differentiate between Nain and Rae Province tonalites in the northern Torngat region, however when coupled with the 3.0 Ga crustal age from the Rae Province, they verify the presence of relatively unreworked Mid- to Late Archean crust within this segment of the Rae Province.

The 2.7 Ga zircon age from these gneisses represents either an igneous or a metamorphic age (D. Scott., pers. comm., 1997). and thus raises the possibility that this part of the Rae Province, like the Nain, was metamorphosed at ca. 2.7 Ga. However, the Sr and Pb systems do not reflect a late Archean LILE depletion event similar to that observed in the adjacent Nain province.

A 2.77 Ga metamorphic zircon age was obtained from a tonalitic gneiss in the Burwell Domain (Connelly and Mengel, 1995) and provides a minimum age for these Archean crustal blocks. However, similar to rocks in the Rae Province, the Burwell gneisses show no evidence for a late Archean LILE depletion event, as is documented in the adjacent Nain Province. The tonalitic samples from the Burwell Domain and Rae Province only rarely preserve relict orthopyroxene (e.g., as cores in hornblende), in contrast to what is commonly observed in Nain Province gneisses in the Four Peaks Domain. This suggests either that Burwell and Rae rocks were more completely retrogressed than Nain Province equivalents, or that metamorphism of these rocks never exceeded upper amphibolite facies conditions, implying that they represent shallower crustal levels relative to the Nain.

4. Among Nain Province gneisses, the $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ are in general positively correlated, indicating that the U-Pb and Th-Pb systems have remained coupled over time. In contrast, $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ values are not similarly correlated among Rae Province samples. The Rae Province tonalites and granites show a strong Th-enrichment which does not

occur in the Nain Province. Tonalites from the Burwell domain are also Th enriched and they have $^{208}Pb/^{204}Pb$ values are similar to those of Rae Province gneisses.

2.7 CONCLUSIONS

Archean gneisses within the Four Peaks domain, northernmost Labrador, have unradiogenic Sr and Pb isotopic signatures which are also a characteristic feature of Nain Province gneisses in the Saglek block further to the south. The Archean gneisses within and immediately west of the KSZ can be linked to the Four Peaks Domain of the Nain Province on the basis of similar lithologies and crystallization ages and the characteristically unradiogenic Pb and Sr isotopic compositions in both areas. In contrast, Pb and Sr isotopic compositions for Archean crustal blocks in the southwestern Burwell domain are strongly radiogenic, and the rocks are geochemically and isotopically much more similar to the adjacent Rae Province gneisses.

The above data shows that the Burwell domain is underlain by Archean basement gneisses which are not allochthonous with respect to the neighboring Nain or Rae provinces during Early Proterozoic convergence and collision. The Burwell domain is in fact a composite terrane, containing elements of both neighboring Archean provinces.

Anorthosites and associated amphibolites with MORB-like compositions occurring in the KSZ are compositionally and lithologically similar to rocks of oceanic affinity formed at a mid-ocean ridge or back arc basin. Consequently, the present location of the anorthosites and amphibolites tectonically juxtaposed with continental crust within the KSZ may mark the location of a former back-arc or rift which was subsequently closed prior to Early Proterozoic calc-alkaline magmatism.

REFERENCES

Arth, J.G., 1979. Some trace elements in trondhjemites-their implications to magma genesis and paleotectonic setting. *In* Trondhjemites, Dacites and Related Rocks, *Ed.* F. Barker, 123-132.

Arth, J.G., Barker, F., 1976. Rare-earth partitioning between hornblende and dacitic liquid and implications for the genesis of trondhjemitic-tonalitic magmas. Geology, 4: 534-536.

Ashwal, L.D., 1993. Anorthosites, Springer-Verlag Publishers, 422pp.

Bardoux, M., David, J., Gariepy, C., Robillard, M., 1996. A progress report on the geology and geochronology of the northeastern Rae Province in southeastern Ungava Bay. *In* ESCOOT Lithoprobe Report 57; 3-14.

Barker, F., Arth, J.G., 1976. Generation of trondhjemitic-tonalitic liquids and Archean bimodal trondhjemitic basalt suites. Geology, 4: 596-600.

Bertrand, J.-M., Roddick, J. C., Van Kranendonk, M. V. & Ermanovics, I. (1993) U-Pb geochronology of deformation and metamorphism across a central transect of the Early Proterozoic Torngat Orogen, North River map area, Labrador. Can. J. Earth Sci. 30: 1470-1489.

Bridgwater, D., Wardle, R.D., 1992. Whole rock Pb isotopic compositions of Archean and Proterozoic rocks from northern Labrador used to separate major units in the shield. *In* ESCOOT Lithoprobe Report 32: 42-51.

Bridgwater, D., Schiotte, L., 1991. The Archean gneiss complex of northern Labrador: A review of current results, ideas and problems. Bull. geol. Soc. Denmark; 39: 153-166.

Bridgwater, D., Collerson, K.D., Hurst, R.W., Jesseau, C.W., 1975. Field characteristics of the early Precambrian rocks from Saglek, coast of Labrador. *In* Report of Activities, part A, Geol. Surv. Canada Paper 75-1: 287-296.

Collerson, K.D., Kerr, A., Vocke, R.D., Hanson, G.N., 1982. Reworking of sialic crust as represented in late Archean-age gneisses, northern Labrador. Geology; 10: 202-208.

Connelly, J. N., Mengel, F. C., 1995. Preliminary U-Pb data from Archean gneisses in a transect across the Komaktorvik Shear Zone, Northern Torngat Orogen. *In* ESCOOT Lithoprobe Report 45; 72-82.

Connelly, J. N., Mengel, F. C., 1996. The Archean backdrop to Paleoproterozoic tectonism in the Nagssugtoqidian and Torngat Orogens: new constraints from U-Pb geochronology. In ECSOOT Lithoprobe Report 57: 15-24.

DePaolo, D.J., 1981. Neodymium isotopes in the Colorado Front Range and crust-mantle evolution in the Proterozoic. Nature, 291: 193-196.

Drummond, M.S., Defant M.J., 1990. A model for trondhjemite-tonalite-dacite genesis and crustal growth via slab melting: Archean to modern comparisions. Jour. Geophys. Res., 95/B13: 503-521.

Ellis, K.J., Thompson, A.B., 1986. Subsolidus and partial melting reactions in the quartz-excess $CaO + MgO + Al_2O_3 + SiO_2 + H_2O$ system under water-excess and water deficient conditions to 10 kb: some implications for the origin of peraluminous melts from mafic rocks. J. Petrol., 27: 91-121.

Ermanovics, I., 1993. Geology of the Hopedale Block, southern Nain Province, and the adjacent Proterozoic terranes, Labrador, Newfoundland. Geological Survey of Canada Memoir 431: 161pp.

Ermanovics, I., Van Kranendonk, M.J., Corriveau, L., Mengel, F., Bridgwater, D., Sherlock, R., 1989. The boundary zone of the Nain-Churchill provinces in the North River-Nutakmap areas, Labrador. Geological Survey of Canada, Paper 89-1C: 385-394.

Evensen, N.M., Hamilton, P.J., O'Nions, R.K., 1978. Rare-earth abundances in chondritic meteorites. Geochim. et Cosmochim. Acta, 42: 1199-1212.

Ewart, A., 1982. The mineralogy and petrology of Tertiary-Recent orogenic volcanic rocks: with special reference to the andesitic-basaltic compositional range. *In* Andesites: orogenic andesites an related rocks. *Ed.* R.S. Thorpe: 26-87.

Gariépy, C., Dupré, B., 1991. Pb isotopes and crust-mantle evolution. *In* Mineralogical Association of Canada- Short course handbook on applications of radiogenic isotope systems to problems in geology. *Eds.* L. Heaman and J. Ludden: vol. 19: 191-224.

Glikson, A.Y., 1979. Early Precambrian tonalite-trondhjemite sialic nuclei. Earth Sci. Rev., 15: 1-73.

Goulet, N. 1992. Structural pattern of the northern Torngat Orogen. *In* ESCOOT Lithoprobe Report 32: 52-53.

Henderson, P., 1984. Rare Earth Element Geochemistry. Elsevier Publishers. 510pp.

Hoffman, P.F., 1990. Dynamics of the tectonic assembly of northeast Laurentia in geon 18 (1.9-1.8 Ga). Geoscience Canada, 17/4: 222-226.

Hoffman, P.F., 1988. United Plates of America, the birth of a craton: Early Proterozoic assembly and growth of Laurentia. Ann. Rev. Earth Planet. Sci., 16: 543-603.

Irvine, T.N., Barager, W.R.A., 1971. A guide to the chemical classification of the common volcanic rocks. Canadian Journal of Earth Sciences, 8: 523-548.

James, D. T., Connelly, J. N., Wasteneys, H. A., Kilfoil, G. J., 1995. Paleoproterozoic lithotectonic divisions of the southeastern Churchill Province, western Labrador. Canadian Journal of Earth Sciences, 33: 216-230.

Kay, R. W., Mahlburg-Kay, S., 1991. Creation and destruction of lower continental crust. Geologische Rundschau, 80:2: 259-278.

Korstgaard, J., Ryan, B., Wardle, R. 1987. The boundary between Proterozoic and Archean crustal blocks in central west Greenland and northern Labrador. In Evolution of the Lewisian and Comparable Precambrian High Grade Terrains. *Eds.* R.G. Park, J. Tarney. Geol. Soc. Sec. Pub. 27: 247-259.

Le Maitre, R.W., (*Ed.*), 1989. A Classification of Igneous Rocks and Glossary of Terms. Blackwell, Oxford. 193 pp.

Lewry, J. F., Collerson, K. D., 1990. The Trans-Hudson Orogen: extent, subdivision, and problems. *In* The Early Proterozoic Trans-Hudson Orogen of North America. *Eds.*J.F. Lewry and M.R. Stauffer: 1-14.

Loveridge, W.D., Ermanovics, I.F., Sullivan, R.W., 1987. U-Pb ages on zircon from the Maggo gneiss, the Kanairiktok plutonic suite and the Island Harbour plutonic suite, coast of Labrador, Newfoundland *In* Radiogenic Age and Isotopic Studies: Report 1. Geol. Surv. Can., Pap., 87-2: 59-65.

Martin, H., 1986. Effect of steeper Archean geothermal gradient on geochemistry of subduction-zone magmas. Geology, 14: 753-756.

Martin, H., 1987. Petrogenesis of Archean Trondhjemites, Tonalites and Granodiorites from Eastern Finland: Major and Trace Element Geochemistry. Jour. Petrol., 28/5: 921-953.

Martin, H., 1993. The mechanisms of petrogenesis of the Archean continental crust-Comparison with modern processes. Lithos, 30: 373-388.

McCulloch, M.T., Gamble, J.A., 1990. Geochemical and geodynamical constraints on subduction zone magmatism. Earth Planet. Sci. Let., 102: 358-374.

Rivers, T., Mengel, F., Scott, D.J., Campbell, L.M., Goulet, N., 1996. Torngat Orogen-a Paleoproterozoic example of a narrow doubly vergent collisional orogen. *In* Precambrian Crustal Evolution in the North Atlantic Region. *Ed.*T.S. Brewer. Geol. Soc. Spec. Pub. 112: 117-136.

Rudnick, R.L., McLennan, S.M., Taylor, S.R., 1985. Large ion lithophile elements in rocks from high-pressure granulite facies terrains. Geochim. et Cosmochim. Acta: 49: 1645-1655.

Schiotte, L., Nutman, A.P., Bridgwater, D., 1991. U-Pb ages of single zircons within "Upernavik" metasedimentary rocks and regional implication for the tectonic evolution of the Archean Nain Province, Labrador. Can. J. Earth Sci., 29: 260-276.

Schiotte, L., Noble, S., Bridgwater, D., 1990. U-Pb mineral ages from northern Labrador: possible evidence for interlayering of early and middle Archean tectonic slices. Geosci. Can., 17: 227-231.

Scott, D.J., 1995a. U-Pb geochronology of a Paleoproterozoic continental magmatic arc on the western margin of the Archean Nain craton, northern Labrador, Canada. Can. J. Earth Sci., 32/11: 1870-1882.

Scott, D.J., 1995b. U-Pb geochronology of the Nain craton on the eastern margin of the Torngat Orogen, Labrador. Can. J. Earth Sci., 32/11: 1859-1869.

Scott, D.J., Machado, N., 1995. U-Pb geochronology of the northern Torngat Orogen, Labrador, Canada: a record of Paleoproterozoic magmatism and deformation. Precambrian Res., 70: 169-190.

Scott, D., Campbell, L.M., 1993. Evolution of the Paleoproterozoic Torngat Orogen, Labrador Canada: Recent advances using U-Pb Geochronology and Nd isotopic systematics. Geological Society of America, Abstracts with Program, vol. 25, p. A-237.

Stacey, J. S., Kramers, J. D., 1975. Approximation of terrestrial lead isotope evolution by a two-stage model. Earth Planet. Sci. Lett., 26: 207-221.

Stockwell, C.H., 1961. Structural provinces, orogenies, and time classification of rocks of the Canadian Precambrian Shield. *In* Age Determinations by the Gological Survey of Canada. *Ed.* J.A. Lowden., Geol. Surv. Can. Pap., 61-17: 108-118.

Streckeisen, A., 1976. To each plutonic rock its proper name. Earth Sci. Rev.: 12: 1-33.

Sun, S. S., 1980. Lead isotopic study of young volcanic rocks from mid-ocean ridges, ocean islands and island arcs. Phil. Trans. R. Soc. Lond.: A297: 409-445.

Taylor, S. R., McLennan, S.M., 1985. The Continental Crust: its Composition and Evolution. Blackwell Scientific Publications, 312pp.

Taylor, F.C., 1979. Reconnaissance geology of a part of the Precambrian Shield in northeastern Quebec, northern Labrador and Northwestern Territories. Geologica Survey of Canada. Memoir 393: 99pp.

Theriault, R.J., Ermanovics, I., 1996. Sm-Nd isotopic and geochemical characterization of the Paleoproterozoic Torngat orogen, Labrador, Canada. Precambrian Res., 81: 15-35.

Touret, J.L.R., 1996. LILE-depletion in granulites: Myth or reality? *In* Petrology and Geochemistry of magmatic suites of rocks in the continental and oceanic crusts. A volume dedicated to Professor Jean Michot *Ed.* D Demaiffe: 53-72.

Van Kranendonk, M.J., Scott, D.J., 1992. Preliminary report on the geology and structural evolution of the Komaktorvik Zone of the Early Proterozoic Torngat Orogen, Eclipse Harbour area, northern Labrador Current Research, Part C: Geological Survey of Canada. Paper 92-1C: 59-68.

Van Kranendonk, M.J., Wardle, R.J., 1993. Geological synthesis and musing on possible subduction-accretion models in the formation of the northern Torngat Orogen. *In* ESCOOT Report of Transect Meeting. *Eds.* R.J. Wardle, J. Hall; Lithoprobe report 34: 62-80.

Van Kranendonk, M.J., Wardle, R.J., 1996. Burwell domain of the Paleoproterozoic Torngat Orogen, northeastern Canada: Tilted cross-section of a magmatic arc caught between a rock and a hard place. *In* Precambrian Crustal Evolution in the North Atlantic Region. *Ed.*T.S. Brewer, Geol. Soc. Spec. Pub. 112: 91-115

Van Kranendonk, M.J., Wardle, R.J., Mengel, F., Campbell, L.M., Reid, L., 1994. New results and summary of the Archean and Paleoproterozoic geology of the Burwell domain, northern Torngat Orogen, Labrador, Quebec, and Northwest Territories. Current Research, Geological Survey of Canada, 1994-C: 321-332.

Wardle, R.J., 1983. Nain-Churchill province cross-section, Nachvak fiord, Northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 83-1: 68-89.

Wardle, R.J., 1984. Nain-Churchill Province cross-section: Riviere Baudancourt-Nachvak Lake. *In* Current Research, Newfoundland Dept. of Mines and Energy, Mineral Development Division, Report 84-1: 1-11.

Wardle, R.J., Bridgwater, D., Mengel, F., Campbell, L., Van Kranendonk, M.J., Hauman, A., Churchill, R., Reid, L., 1994. Mapping in the Torngat Orogen, northernmost Labrador: Report 3, the Nain caton (including a note on ultramafic dyke occurrences in northernmost Labrador). *In* Current Research, Newfoundland Dept. of Mines and Energy, Geol. Surv. Branch, Report 94-1: 399-407.

Wardle, R.J., Van Kranendonk, M.J., Mengel, F., Scott, D., 1992. Geological mapping in the Torngat Orogen, northernmost Labrador: Preliminary results. Current Research., Newfoundland Dept. of Mines and Energy, Geological Survey Branch, Rept. 92-1: 413-429.

Wardle, R.J., Van Kranendonk, M.J., Mengel, F., Scott, D., Schwarz, S., Ryan, B., Bridgwater, D., 1993. Geological mapping in the Torngat Orogen, northernmost Labrador: Report 2. *In* Current Research, Newfoundland Dept. of Mines and Energy, Geol. Surv. Branch, Report 93-1: 77-89.

Wilson, M., 1989. Igneous Petrogenesis. Unwin Hyman Pubs. 466pp.

Winther, T.K., Newton, R.C., 1991. Experimental metling of a hydrous low-K tholeiite: evidence on the origin of Archean Cratons. Bull. Geol. Soc. Den., 39: 213-228.

CHAPTER 3

Isotopic and Geochemical constraints on Petrogenesis of Early Proterozoic Igneous Rocks in the northern Torngat Orogen, Labrador: Implications for Tectonic Assembly of the Torngat Orogen

3.1 INTRODUCTION

One of the fundamental questions about the geological and geochemical evolution of the Earth involves the tectonic and petrologic mechanisms responsible for formation of continental crust. During the Phanerozoic, it is apparent that new continental crust was generated at convergent plate margins through partial melting of the mantle wedge above subducted oceanic lithosphere. In this setting, melting is triggered by fluids added to the wedge through dehydration of the oceanic lithosphere itself (e.g., Wyllie and Sekine, 1982; Tatsumi, 1989). In the Precambrian, however, this process is more difficult to assess because Precambrian crustal terranes are typically eroded to very deep levels and have been subjected to multiple periods of tectonism and metamorphism. As a result, it is more difficult to extract information regarding the original processes involved in the formation and stabilization of such ancient crustal segments, despite the importance of such information for understanding how early continental crust was formed and for assessing the importance of Phanerozoic-style plate tectonics in the Precambrian.

In this paper, we document the major and trace element geochemistry and Nd, Sr and Pb isotopic composition of metaigneous rocks which were added to the Archean crust of northern Labrador during development of the Early Proterozoic Torngat Orogen (Figs. 1 and 2).

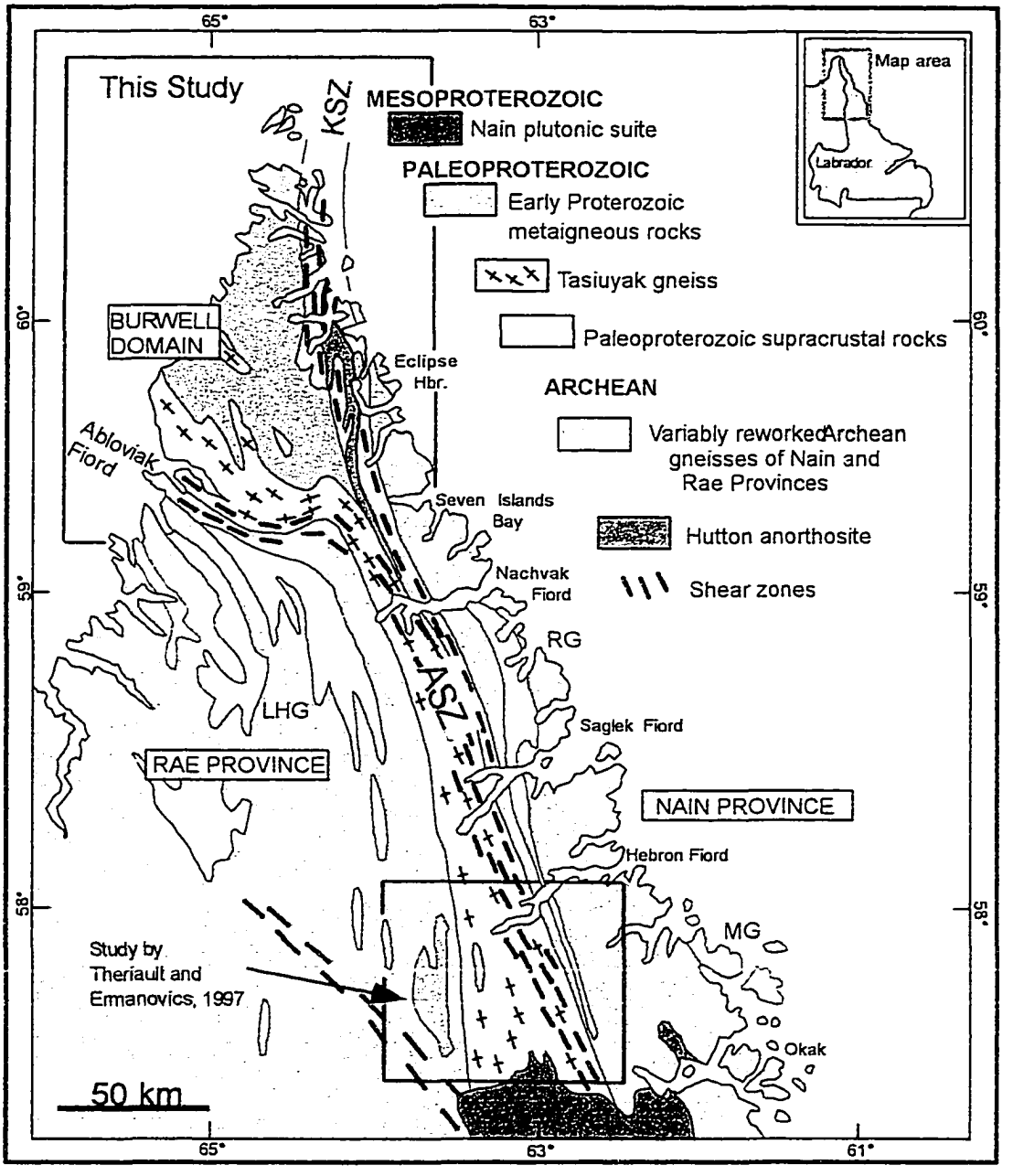


Figure 1. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group. RG = Ramah Group. MG = Mugford group. Also shown are the location of Figure 2 ("This study") and the area described by Theriault & Ermanovics (1997).

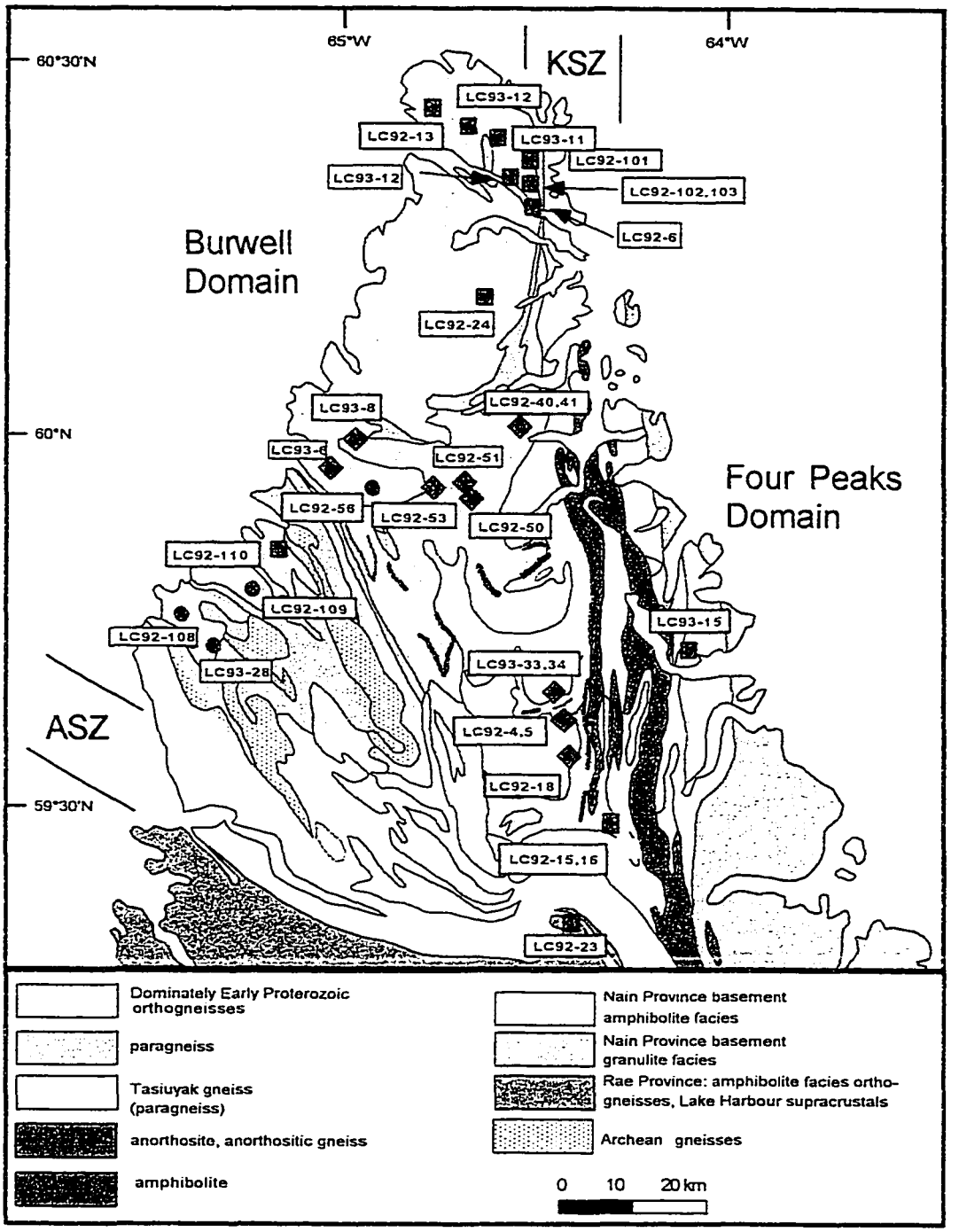


Figure 2. Simplified geological map of the study area in northernmost Labrador (modified from Wardle et al., 1993: VanKranendonk et al., 1994). For location, see Figure 1. ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively (see also Figure 1). Sample locations and distribution of members of Groups 1, 2 and 3 are also shown (see text for description of Groups). Symbols are as follows: samples in Group 1 are shown as diamonds; Group 2 as squares, Group 3 as circles.

The purpose of this work is to: (1) constrain the processes by which these Early Proterozoic arc rocks formed, and compare these to Phanerozoic analogs; and (2) provide constraints on the tectonic evolution of the Torngat Orogen by identifying and characterizing the spatial distribution of isotopic and geochemical signatures of the Proterozoic igneous rocks.

The Torngat Orogen is well-suited for this study because its structural evolution is well-constrained and a good geochronological database is available (see below). In addition, critical field relationships such as intrusive contacts between the Proterozoic magmatic rocks and the Archean basement gneisses are well exposed and easily inspected.

The Archean basement gneisses, into which Proterozoic magmas are emplaced, are described in a companion paper (Chapter 2). These gneisses are isotopically distinct from the younger magmas, and crustal assimilation is therefore easy to identify on the basis of initial Nd isotopic composition.

3.2 REGIONAL GEOLOGY OF THE TORNGAT OROGEN

The Torngat Orogen (TO) developed as the result of transpressional collision in the period 1860-1770 Ma between two Archean crustal blocks, the Nain Province to the east, and the Rae Province to the west (e.g., Korstgaard et al., 1987; Mengel, 1988; Van Kranendonk et al., 1994, Van Kranendonk and Wardle, 1997; Fig. 1). The suture-zone between the two Archean crustal blocks is overlain by a thick belt of metasediments (Tasiuyak gneiss; Wardle, 1983), which coincides with, and is affected by, the Abloviak shear zone (ASZ), a crustal scale, oblique-sinistral shear zone that was active in the period ca. 1845-1822 Ma (Scott & Machado, 1994, 1995; Bertrand et al., 1993). Early Proterozoic igneous rocks occur sporadically throughout the TO. In the southern segment, i.e., at the latitude of Okak (Fig. 1), these rocks occur to the west of and within

the core of the orogen (e.g., the Tasiuyak gneiss and Abloviak shear zone) (Fig. 1: Theriault and Ermanovics, 1996), while in the northern segment they occur in the eastern marginal zone of the orogen, either as intrusive bodies into Nain Province crust or as sheets tectonically intercalated with Archean basement gneisses and supracrustal rocks. West of the orogen, Rae Province gneisses are generally pervasively reworked and migmatized (Van Kranendonk & Ermanovics, 1990; Ermanovics & Van Kranendonk, 1990), whereas to the east, Nain Province gneisses are only moderately affected by Proterozoic deformation (Wardle et al., 1992). The orogen is presently exposed at mid- to lower crustal depths as shown by the occurrence of amphibolite to granulite facies mineral assemblages in all lithologies.

3.2.1 The Northern Torngat Orogen

In the northern-most segment of the orogen, the north-trending Abloviak shear zone and Tasiuyak gneiss are deflected to the west, and the younger Komaktorvik shear zone (KSZ; active from 1798-1774 Ma; Scott & Machado, 1994, 1995), extends to the north (Fig. 1). Within the Komaktorvik shear zone, the Hutton anorthosite and associated amphibolites form a linear belt that extends southward for more than 100 km (Figs. 1 and 2). The crustal block between the Abloviak and Komaktorvik shear zones is called the Burwell domain (Korstgaard et al., 1987; Van Kranendonk et al., 1992), and represents an enigmatic collage of tectonically juxtaposed para- and orthogneisses, many of which are suspected to be Archean on the basis of their polycyclic appearance and the presence of mafic dykes (Fig. 2). The linkage between the presumed Archean gneisses in the Burwell domain and those in the adjacent Nain and Rae Provinces was discussed in Chapter 2. Mafic dykes ranging in age between ca. 2.4 - 2.2 Ga intrude the Archean gneisses on either side of the KSZ (Van Kranendonk and Wardle, 1997; Connelly and Mengel, 1994; Ermanovics, 1993; Taylor, 1974).

Early Proterozoic metaigneous rocks which occur in the northern TO vary from east to west in terms of the metamorphic overprint and contact relationships to host rocks. In the western part of the Burwell domain, they typically occur as tectonic slices and intrusive sheets, although contacts are often obscured. They are generally retrogressed from granulite to amphibolite facies, as shown by amphibole grains which variably rim or replace orthopyroxene. Their igneous ages range between 1895 and 1869 Ma (e.g., Scott, 1995a). Comparable metaigneous rocks occurring in the eastern part of the Burwell domain and east of the KSZ have igneous ages of 1910-1864 Ma (Scott 1995a). Here they commonly occur as homogeneous magmatic bodies intruding the structurally more complex host gneisses. In general they are mildly deformed and commonly preserve orthopyroxene, which is usually rimmed by hornblende.

3.2.2 Current models for development of Torngat Orogen

Previous studies in the Torngat region have interpreted the Early Proterozoic igneous rocks as segments of a continental arc which developed above an Early Proterozoic subduction zone. Studies focusing on the southern segment of the orogen proposed a westward dipping subduction zone to explain the geographic position of Early Proterozoic magmatic rocks within the Rae Province to the west (Fig. 1; Theriault and Ermanovics, 1996; Bertrand et al., 1993; Van Kranendonk and Ermanovics, 1990). In contrast, studies on the northern segment of the orogen have invoked eastward dipping subduction to explain the concentration of Early Proterozoic magmatism in the eastern part of the orogen and in the Nain Province foreland (Scott, 1995; Scott and Campbell, 1993; Rivers et al., 1996). In an attempt to integrate all observations, a complex model proposing dual subduction in the northern segment of the TO has also been introduced (Van Kranendonk and Wardle; *in press*).

3.3 SCOPE OF THIS STUDY

Prior to this investigation, no studies have addressed the petrogenesis of the Early Proterozoic igneous rocks, evaluated them within the framework of better understood Phanerozoic subduction zone processes or documented lateral variations in the isotopic and geochemical characteristics of these rocks in support of the proposed subduction zone geometries. These themes represent the basis of this study.

In order to address these problems, 32 Early Proterozoic metaigneous rocks from throughout the northern Torngat region were sampled (Fig. 2), and analyzed for major and trace element concentrations. A subset of these samples was selected for Sm-Nd, Rb-Sr and Pb-Pb isotope analysis. All samples were selected from areas of low strain to minimize effects of post-emplacement deformation and metamorphism.

The similarity in crystallization ages and the spatial continuity of the western and eastern rocks would suggest that they formed as part of the same arc suite, but were affected differently by deformation and metamorphism which was heterogeneous throughout the orogen. However, it is not certain whether the Early Proterozoic metaigneous rocks in the east were derived from the same sources as those in the west: therefore these rocks have been subdivided on the basis of their initial Nd isotopic compositions (shown here as $\varepsilon Nd_{(1895)}$, using 1895 Ma as the approximate crystallization age, see above). The rationale here is that the $\varepsilon Nd_{(1895)}$ will reflect the source(s) for the rocks, and thus ensure that comparisons are made between rocks with similar source characteristics. The Early Proterozoic metaigneous rocks were subdivided into three groups: Group 1: $\varepsilon Nd_{(1895)} > 0$, Group 2: $-5 < \varepsilon Nd_{(1895)} < 0$, and Group 3: $\varepsilon Nd_{(1895)} < -5$. The geographic distribution of the groups is shown in Figs. 2 and 3

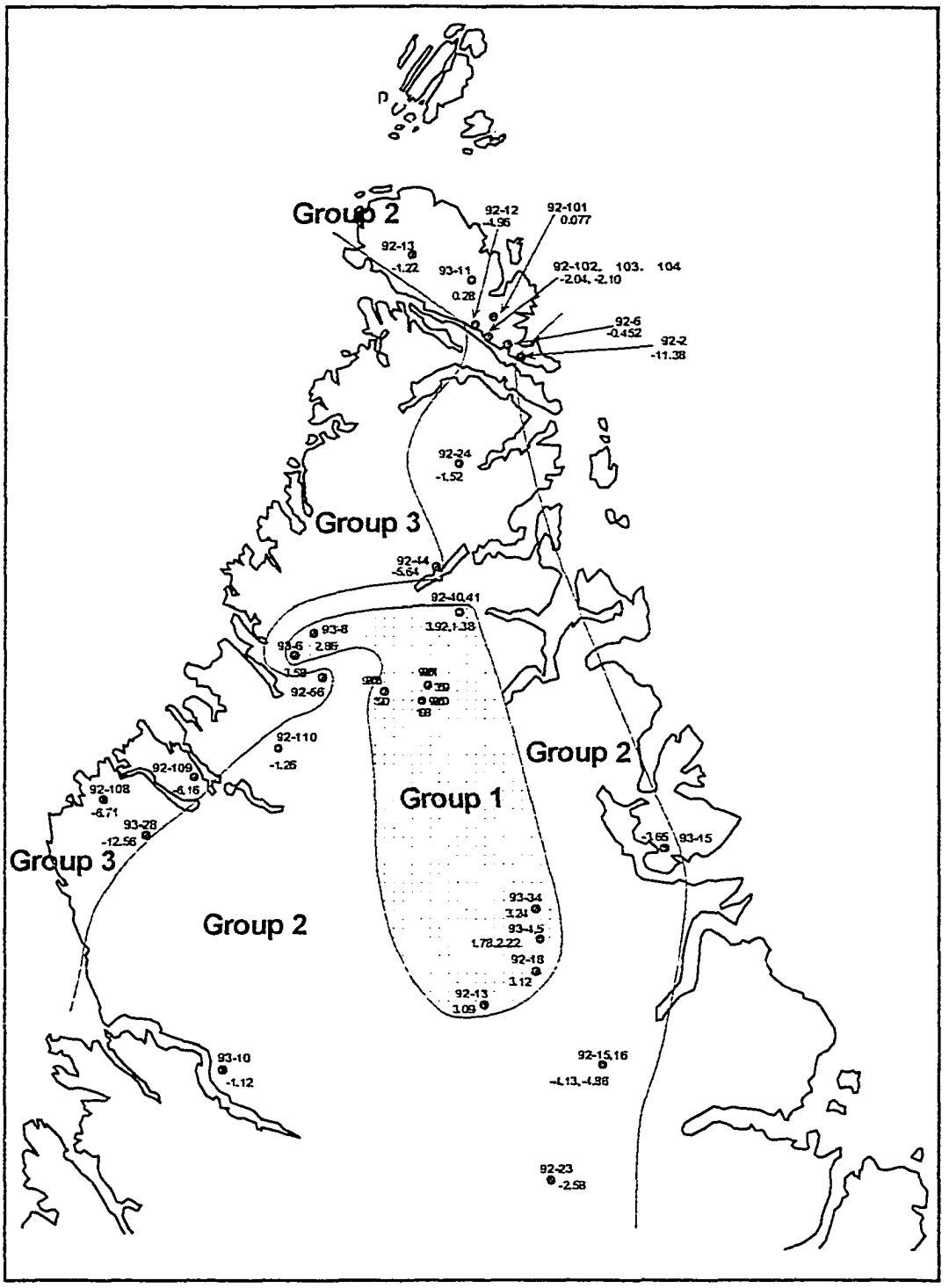


Figure 3. Geographic distribution of $\varepsilon_{Nd\,(1895)}$ values for Early Proterozoic metaigneous rocks in the Torngat Orogen, Northern Labrador. Basemap identical to Figure 2. Contours based on values of $\varepsilon_{Nd\,(1895)}$, i.e., $\varepsilon_{Nd\,(1895)} > 0$ (Group 1), $-5 < \varepsilon_{Nd\,(1895)} < 0$ (Group 2) and $\varepsilon_{Nd\,(1895)} < -5$ (Group 3).

3.4 DESCRIPTION OF SAMPLES

3.4.1 Group 1 ($\varepsilon Nd_{(1895)} > 0$)

3.4.1.1 Field Description

The fifteen samples which comprise Group 1 occur along a N-S axis in the center of the study area (Figs. 2 and 3). In the field, Group 1 rocks are generally homogeneous, often massive or preserve evidence for a single phase of deformation. They may also have sharp intrusive or tectonic contacts with lithologically and structurally more complex host gneisses. At several locations, Group 1 rocks intrude the Archean gneisses, but are also cut by a later phase of Group 1 intrusives. For example, a gabbro cumulate (LC92-50) intrudes Archean Nain Province gneisses, but is itself intruded by a tonalite (LC92-51).

3.4.1.2 Petrography

Most of the samples in Group 1 are partially or completely recrystallized during granulite facies metamorphism, and only a few preserve recognizable igneous textures. In general, the samples are weakly foliated to isotropic. Quartz is strongly strained and commonly shows domains of incipient recrystallization. Orthopyroxene (< 10%), is typically rimmed by hornblende. In some samples the orthopyroxene is entirely replaced by hornblende. Several samples preserve orthopyroxene as symplectic intergrowths with plagioclase, suggesting that an earlier garnet-clinopyroxene assemblage was replaced by orthopyroxene-plagioclase during decompression following peak metamorphism.

Mafic rocks of Group 1 are dominated by orthopyroxene, clinopyroxene, plagioclase and hornblende. Pyroxenes are commonly rimmed, and variably replaced by hornblende. Accessory phases include apatite, sphene and Fe-Ti-oxides. In one sample (LC92-40), sphene constitutes approximately 5% of the rock. Minor biotite replaces hornblende in some rocks. More felsic samples are composed of plagioclase (< 60%),

quartz (10-30%), hornblende and/or biotite (< 20%) and minor K-feldspar (0-5%). More strongly retrogressed samples contain minor chlorite replacing biotite. Accessory phases in the felsic rocks are zoisite, apatite and zircon.

3.4.2 Group 2 ($\varepsilon Nd_{(1895)} = -5$ to 0)

3.4.2.1 Field Description

Group 2 samples (n = 13) occur in the eastern, western and northern parts of the northern TO (Figs. 2 and 3) as tectonic slices and as intrusive sheets and bodies in Archean basement gneisses and metasediments (Fig. 2). Samples from Group 2 exhibit a range of textures ranging from primary igneous with a weak foliation defined by mafic minerals to well developed metamorphic fabrics including mineral foliation and ompositional layering defined by mafic and felsic minerals. Some of the samples show thorough granoblastic recrystallization which has obscured any previous fabrics.

3.4.2.2 Petrography

Pyroxenes are less abundant in Group 2 samples than in Group 1, and comprise < 10% orthopyroxene and < 5% clinopyroxene. In all cases pyroxenes are replaced by hornblende and biotite, particularly along grain boundaries and cracks. Biotite is common and is typically reddish. Plagioclase contents range from 30-50%, and many of the samples also contain microcline (< 5-30%). Quartz ranges from < 10-40% and zoisite may be abundant (up to ca. 30% in LC92-103). Minor calcite and garnet occur in some samples. Accessory minerals (< 5%) include apatite, zircon and Fe-Ti-oxides.

3.4.3 Group 3 ($\varepsilon Nd_{(1895)} < -5$)

3.4.3.1 Field Description

The four Group 3 samples were collected from the western-most part of the Burwell domain (Figs. 2 and 3). In this area these rocks are deformed, migmatized and tectonically intercalated with host gneisses, such that boundaries with adjacent lithologies are obscured.

3.4.3.2 Petrography

The metaigneous rocks comprising Group 3 are petrographically distinct from samples in Groups 1 and 2. Group 3 rocks show minimal granoblastic recrystallization and in general preserve igneous textures with irregular, rounded grain boundaries. The samples contain irregularly shaped garnet grains and rounded zircons, which, on the basis of their morphology, may represent xenocrysts. Quartz is strained and contains rutile needles. The samples contain minor orthopyroxene which is largely altered to biotite and chlorite. Biotite grains are commonly kinked. Plagioclase is abundant (20-30%), and sometimes strained. All samples contain microcline (ca. 20%). Accessory phases include chlorite, apatite and muscovite.

3.5 PRESENTATION OF DATA

3.5.1 Major And Trace Element Geochemistry

Abundances of major elements (recalculated to 100% anhydrous) and trace elements for 32 samples of Early Proterozoic metaigneous rocks from the northern-most Torngat Orogen are summarized in Tables 1 and 2; sample locations are shown in Fig. 2.

TABLE 1 - Major element analyses of Early Proterozoic metaigneous rocks, Torngat Orogen, Northern Labrador

	GROUP 1										
(Wt.%)	LC92-018	LC92-041	LC92-051	LC92-053	LC92-101	LC93-004	LC93-005	LC93-006	LC93-008	LC93-011	LC93-034
SiO2	68.9	65.7	71.0	51.0	59.3	72.0	70.7	56.7	59.6	56.5	72.5
TiO2	0.21	0.58	0.21	1.02	0.75	0.14	0.12	0.73	0.43	0.97	0.21
Al2O3	16.94	15.91	16.11	20.96	16.92	16.05	16.70	17.15	19.36	17.45	14.47
Fe2O3T	2.43	4.43	2.13	8.73	7.15	1.40	1.39	8.87	5.09	9.37	1.81
MnO	0.04	0.06	0.03	0.11	0.11	0.02	0.03	0.12	0.07	0.14	0.03
MgO	1.12	2.17	0.83	3.15	3.08	0.68	0.63	3.97	3.08	3.29	1.22
CaO	3.58	4.25	3.55	8.98	6.20	3.25	3.01	8.44	7.44	7.20	2.53
Na2O	4.97	4.53	4.86	4.71	3.52	5.08	5.27	3.59	4.19	3.49	4.92
K2O	1.35	1.55	0.94	0.73	2.13	0.94	1.47	0.16	0.19	1.24	1.21
P2O5	0.07	0.18	0.07	0.25	0.19	0.10	0.10	0.20	0.20	0.30	0.10
H2O(T)	0.40	0.70	0.20	0.50	0.60	0.30	0.40	0.40	0.50	0.40	0.70
CO2(T)	-0.10	0.10	-0.10	0.20	0.20	0.10	0.20	0.10	0.20	0.10	0.40
Sum	98.60	99.30	98.70	99.70	99.30	100.30	100.60	100.30	100.20	100.30	99.50

	GROUP 1 (tholeiites)				GROUP 2						
(Wt.%)	LC92-013	LC93-033	LC92-040	LC92-050	LC92-006B	LC92-012	LC92-016	LC92-023	LC92-024	LC92-102	LC92-103
SiO2	48.8	46.1	51.3	52.2	59.0	60.5	56.9	64.8	62.5	60.3	53.5
TiO2	1.10	0.89	1.82	0.25	0.83	0.74	0.97	0.61	0.73	0.70	1.41
Al2O3	13.50	15.38	15.30	12.93	17.35	19.01	16.30	14.99	17.59	18.71	21.17
Fe2O3T	16.30	13.19	13.10	10.02	7.62	3.62	7.50	5.79	5.03	4.63	7.06
MnO	0.24	0.17	0.20	0.18	0.10	0.07	0.10	0.07	0.07	0.06	0.10
MgO	5.58	8.16	4.08	11.02	3.11	0.97	4.30	4.24	0.73	0.74	2.27
CaO	10.50	11.21	8.97	11.12	6.34	3.00	5.85	4.03	3.57	2.52	7.27
Na2O	3.00	2.48	3.50	2.10	3.51	3.72	3.70	2.50	3.62	3.52	5.14
K2O	0.29	0.41	1.20	0.10	1.72	7.37	2.89	2.84	5.50	8.05	0.91
P2O5	0.11	0.10	0.28	0.02	0.21	0.25	0.51	0.15	0.21	0.24	0.59
H2O(T)	1.20	2.50	1.00	0.70	0.50	0.30	1.00	0.20	0.20	0.40	0.30
CO2(T)	0.40	0.20	-0.10	-0.10	0.10	0.10	0.10	0.10	0.10	0.20	0.30
Sum	100.00	100.80	100.00	99.80	99.70	99.40	100.00	100.10	99.50	99.40	99.20

Major element data by X-ray fluorescence except Fe2O3 by ICP.
Oxides normalized to 100% (anhydrous). Actual analysis totals are reported in "Sum" column.
Total iron reported as Fe2O3T

TABLE 1 - (continued)

	GROUP 2						GROUP 3			
(Wt.%)	LC92-015	LC92-110	LC93-010	LC93-013	LC93-015	LC93-016	LC92-056	LC92-108	LC92-109	LC93-028
SiO2	72.8	57.5	58.7	66.5	57.7	56.2	69.3	65.4	71.6	70.6
TiO2	0.26	0.88	0.89	1.14	1.38	0.68	0.30	0.76	0.33	0.35
Al2O3	14.37	18.18	17.88	13.90	16.22	16.83	16.70	17.19	15.11	15.10
Fe2O3T	1.78	7.68	6.79	6.16	7.96	6.83	2.23	3.62	1.32	2.50
MnO	0.02	0.09	0.08	0.08	0.10	0.10	0.01	0.05	0.01	0.02
MgO	0.47	3.25	4.14	1.66	2.68	4.97	0.90	1.50	0.72	1.19
CaO	1.69	6.75	6.87	3.19	5.43	7.37	4.10	4.33	2.02	2.83
Na2O	3.47	3.43	3.40	2.98	3.88	3.66	4.66	5.13	2.94	5.80
K2O	4.46	1.53	0.92	3.08	3.17	1.77	1.12	1.34	5.25	0.78
P2O5	0.07	0.25	0.30	0.30	0.80	0.50	0.07	0.36	0.05	0.20
H2O(T)	0.40	0.70	0.20	0.50	0.80	1.10	0.40	0.30	0.40	0.50
CO2(T)	0.10	0.20	0.20	0.50	0.10	0.20	0.20	0.10	-0.10	0.20
Sum	100.90	99.00	100.10	100.70	100.50	101.00	98.80	99.50	98.60	100.00

Major element data by X-ray fluorescence except Fe2O3 by ICP.
Oxides normalized to 100% (anhydrous). Actual totals are reported in "Sum" column.
Total iron reported as Fe2O3T.

TABLE 2 - Trace element analyses of Early Proterozoic metaigneous rocks, Torngat Orogen, Northern Labrador

	GROUP 1										
	LC92-018	LC92-041	LC92-051	LC92-053	LC92-101	LC93-004	LC93-005	LC93-006	LC93-008	LC93-011	LC93-034
Ba (ppm)	220	420	310	120	850	270	460	60	80	620	280
Co	8.0	18.0	7.0	28.0	27.0						
Cr	<10	51	<10	20	44						
Cs						0.22	0.45	<0.02	<0.02	0.05	0.69
Cu	19	23	<10	73	31						
Ga	20	21	18	23	20	18	21	19	18	22	16
Hf	3.20	3.80	2.30	1.30	4.30	2.60	2.60	2.00	2.10	3.00	2.90
In		0.02	<0.02	0.10	0.05	<0.05	<0.05	0.11	0.09	0.06	<0.05
Nb	1.0	6.0	2.4	4.3	9.6	1.0	3.1	3.1	2.1	9.0	4.3
Ni	<10	33	<10	34	25						
Pb	<20	31	<20	<20	<20	44	11	2	3	10	3
Rb	64.0	57.0	33.0	<10	67.0	26.0	45.0	<10	<10	21.0	35.0
Sc	3.5	8.1	2.9	12.0	18.0						
Sr	890	670	580	910	490	830	730	460	500	360	390
Ta	0.02	0.30	0.10	0.19	0.59	<0.2	<0.2	0.35	<0.2	0.27	0.22
Th	1.6	2.8	0.3	0.5	1.7	1.3	0.8	0.1	0.1	0.4	1.4
Tl	0.38	0.39	0.17	0.04	0.43	0.19	0.28	<0.02	<0.02	0.09	0.17
U	0.32	1.30	0.06	0.38	0.39	0.31	0.27	0.03	0.02	0.15	0.57
V	9	56	12	140	100						
Y	3.6	11.0	3.0	15.0	27.0	2.5	2.7	12.0	6.4	29.0	5.0
Zn	54.0	74.0	41.0	72.0	90.0						
Zr	140.0	150.0	90.0	44.0	180.0	110.0	100.0	72.0	120.0	130.0	130.0
La	12.0	20.0	9.3	9.1	29.0	5.8	5.1	3.8	6.3	26.0	10.0
Ce	22.0	44.0	19.0	23.0	60.0	13.0	11.0	10.0	14.0	57.0	21.0
Pr						1.2	1.1	1.4	1.7	6.4	2.2
Nd	9.9	21.0	7.9	16.0	29.0	4.6	4.4	7.1	7.8	29.0	9.1
Sm	1.70	3.80	1.30	3.70	5.90	0.77	0.97	1.80	1.50	6.00	1.60
Eu	0.53	1.10	0.45	1.50	1.50	0.32	0.29	0.74	0.65	1.70	0.46
Gd	1.10	2.80	0.89	3.40	5.30	0.56	0.78	2.20	1.40	5.90	1.20
Tb											
Dy	0.62	2.00	0.53	2.70	4.50	0.40	0.52	2.00	1.10	5.00	0.87
Ho	0.11	0.37	0.09	0.51	0.93	0.08	0.10	0.43	0.22	1.00	0.17
Er	0.33	0.97	0.23	1.40	2.60	0.22	0.23	1.20	0.61	2.80	0.45
Tm	0.04	0.14	0.03	0.20	0.41	0.04	0.04	0.20	0.10	0.47	0.08
Yb	0.29	0.91	0.17	1.20	2.50	0.23	0.22	1.20	0.60	2.80	0.46
Lu						0.04	0.04	0.18	0.09	0.42	0.08

Analytical details are given in Appendix 1.

TABLE 2 - (continued)

	GROUP 1 (tholeiites)				GROUP 2						
	LC92-013	LC93-033	LC92-040	LC92-050	LC92-006B	LC92-012	LC92-015	LC92-016	LC92-023	LC92-024	LC92-102
Ba (ppm)	100	110	260	60	910	4800	1100	1300	760	3000	1600
Co	58.0		45.0	52.0	26.0	9.0			25.0	12.0	9.0
Cr	50		33	190	52	<10			330	<10	<10
Cs		0.09					0.62	0.60			
Cu	140		80	<10	16	<10			<10	<10	<10
Ga	17	15	21	14	21	17	16	18	19	22	17
Hf	2.40	1.40	4.60	0.67	5.40	2.90	4.50	4.30	4.90	17.00	2.60
In	0.18	0.06	0.08		0.12	0.10	0.06	<0.05	0.04	0.05	<0.02
Nb	2.7	2.1	8.9	0.2	8.0	10.0	8.1	15.0	6.7	23.0	5.4
Ni	55		36	80	28	24			52	17	<10
Pb	<20	4	<20	<20	<20	20	21	11	<20	21	<20
Rb	10.0	19.0	21.0	13.0	63.0	88.0	160.0	100.0	100.0	79.0	84.0
Sc	49.0		35.0	46.0	18.0	11.0			14.0	10.0	5.9
Sr	84	180	300	310	490	580	300	980	320	440	240
Ta	0.19	0.61	0.43	<0.02	0.38	0.47	0.30	1.10	0.20	0.58	0.33
Th	0.2	0.2	2.5	0.1	0.7	0.2	16.0	5.9	2.6	0.6	0.6
Tl	0.07	0.07	0.15	<0.02	0.35	0.47	0.93	0.65	0.61	0.41	0.51
U	0.15	0.07	1.30	0.15	0.48	0.14	0.74	1.10	0.52	0.42	0.26
V	310		260	160	110	18			75	9	<5
Y	38.0	23.0	44.0	6.1	19.0	6.6	3.7	13.0	8.5	21.0	9.2
Zn	150.0		110.0	75.0	92.0	39.0			89.0	77.0	54.0
Zr	88.0	42.0	170.0	24.0	230.0	160.0	210.0	260.0	180.0	730.0	150.0
La	2.6	1.9	17.0	1.1	20.0	27.0	14.0	65.0	35.0	41.0	24.0
Ce	7.3	5.4	40.0	2.6	42.0	49.0	46.0	130.0	68.0	77.0	46.0
Pr		0.8					2.7	15.0			
Nd	6.5	5.3	25.0	2.2	22.0	21.0	8.7	50.0	28.0	38.0	24.0
Sm	2.60	1.90	6.40	0.79	4.50	3.00	1.50	7.30	4.10	6.70	3.90
Eu	1.00	0.86	2.00	0.33	1.50	3.10	0.48	1.90	1.20	3.50	5.70
Gd	4.20	3.10	7.40	0.93	4.00	2.20	0.93	4.40	2.80	5.30	2.80
Tb											
Dy	5.50	3.70	7.40	1.00	3.20	1.20	0.62	2.60	1.50	3.80	1.70
Ho	1.30	0.84	1.50	0.21	0.63	0.22	0.12	0.43	0.29	0.76	0.32
Er	4.00	2.40	4.40	0.62	1.70	0.58	0.33	1.00	0.71	2.00	0.80
Tm	0.66	0.41	0.70	0.10	0.27	0.08	0.05	0.15	0.11	0.31	0.12
Yb	4.10	2.50	4.20	0.60	1.60	0.48	0.34	0.97	0.74	1.90	0.66
Lu		0.39					0.05	0.14			

Analytical details are given in Appendix 1.

TABLE 2 - (continued)

	GROUP 2						GROUP 3			
	LC92-103	LC92-110	LC93-010	LC93-013	LC93-015	LC93-016	LC93-028	LC92-056	LC92-108	LC92-109
Ba (ppm)	660	770	570	2300	1600	890	340	630	400	2000
Co	23.0	24.0						10.0	15.0	5.0
Cr	56	51						11	17	<10
Cs			<0.02	0.13	1.90	0.75	0.03			
Cu	25	16						<10	15	<10
Ga	25	22	23	19	22	18	18	19	21	17
Hf	13.00	5.70	2.10	23.00	9.20	4.30	3.00	3.60	1.10	5.00
In	0.03	0.13	<0.05	0.05	0.05	<0.05	<0.05	0.06	<0.02	0.06
Nb	16.0	9.5	4.3	16.0	36.0	12.0	2.9	16.0	16.0	2.0
Ni	26	28						<20	<20	<10
Pb	<20	<20	6	16	18	15	4	24	12	<20
Rb	15.0	62.0	20.0	65.0	100.0	52.0	19.0	2.3	4.9	140.0
Sc	6.8	27.0						610.0	540.0	2.1
Sr	1200	660	860	370	790	1100	540	0	0	350
Ta	0.55	0.24	<0.2	0.44	1.00	0.55	<0.2	0.08	1.30	0.21
Th	0.9	0.5	0.1	2.9	9.2	5.4	6.8	0.1	0.1	1.1
Tl	0.03	0.36	0.07	0.35	0.79	0.44	0.07	0.06	0.23	0.92
U	0.38	0.29	0.04	0.59	1.90	1.10	0.24	17.00	40.00	0.61
V	79	110						2	5	10
Y	13.0	28.0	9.4	14.0	22.0	11.0	4.9	34.0	78.0	1.3
Zn	88.0	100.0						150.0	51.0	<5
Zr	630.0	240.0	120.0	1000	450.0	220.0	150.0			200.0
								13.0	33.0	
La	63.0	20.0	9.5	53.0	83.0	58.0	74.0	24.0	70.0	15.0
Ce	110.0	50.0	24.0	100.0	170.0	110.0	150.0			24.0
Pr			3.1	11.0	19.0	12.0	16.0	7.9	33.0	
Nd	45.0	32.0	16.0	43.0	73.0	45.0	61.0	1.0	5.2	7.0
Sm	6.20	7.60	3.40	6.10	11.00	6.00	6.40	0.78	0.98	0.84
Eu	2.00	1.70	1.20	1.80	2.40	1.60	1.30	0.63	3.00	0.97
Gd	4.00	6.80	2.80	4.30	7.00	3.70	2.70			0.51
Tb								0.34	1.20	
Dy	2.30	5.00	1.80	2.40	4.40	2.20	1.10	0.06	0.17	0.21
Ho	0.42	0.93	0.35	0.51	0.79	0.39	0.18	0.15	0.35	0.04
Er	1.10	2.30	0.87	1.50	1.90	0.93	0.36	<0.02	0.04	0.10
Tm	0.15	0.36	0.14	0.26	0.31	0.16	0.06	0.11	0.21	<0.02
Yb	0.91	2.00	0.79	1.80	1.70	0.81	0.32			0.11
Lu			0.12	0.32	0.27	0.13	0.06			

Analytical details are given in Appendix 1

3.5.1.1 Geochemistry of Group 1 samples ($\varepsilon Nd_{(1895)} > 0$)

All Group 1 rocks occur in the eastern part of the Burwell domain (Figs. 2 and 3). Based on their normative mineralogy, they are tonalites, quartz diorites and trondhjemites (Figs. 4B and C). The group includes both calc-alkaline (SiO_2 = 50.8-72.2 wt.%) and tholeiitic (SiO_2 = 46.5- 52.1 wt %) rocks (Fig. 3A). Al_2O_3 ranges from 13 to 21 wt.% and is highest at ca. 58-60 wt.% SiO_2 (Fig. 5). MgO, TiO_2, Fe_2O_3, CaO and Y decrease systematically with increasing SiO_2 and Sr and Na_2O increase with increasing SiO_2 (Figs. 5, 6 and 7). All Group 1 samples have low concentrations of K_2O (< 1.6 wt.%), Ba (< 620 ppm), and Nb (< 9 ppm), and these concentrations do not vary with increasing SiO_2. Rb/Sr and K/Rb ratios are also low (< 0.18 and < 500, respectively), and do not change with increasing SiO_2.

The calc-alkaline rocks in Group 1 have moderately fractionated REEs ($La_{(N)}/Yb_{(N)}$ is < 20; and $La_{(N)}/Yb_{(N)}$ increases only slightly with increasing SiO_2 (Fig 7A)). In Figs. 8A and B, these samples are subdivided on the basis of LREE-enrichment, with $La_{(N)}$ = 40-110 and 15-30, respectively (for Figs. 8A and B). In contrast, tholeiitic samples in Group 1 (Fig. 8C), have REE-patterns with flat to slightly negative slopes with lower LREE-abundances, and generally higher HREE-abundances. All Group 1 samples either do not have or have only a small positive Eu anomaly (Fig. 7E, 8A-C).

In N-MORB-normalized spider diagrams (Figs. 9A-C), samples are subdivided on the basis of their REE patterns as in Fig. 8. Both tholeiitic and calc-alkaline members of Group 1 are enriched in Rb, Ba, Pb and K relative to N-MORB, and in general they show a systematic decrease in element enrichment from left to right in these diagrams. For the elements U through Yb, the tholeiitic samples have MORB-like concentrations whereas calc-alkaline samples show decreasing abundances towards the right. The elements are ordered according to decreasing mobility in subduction zone fluids

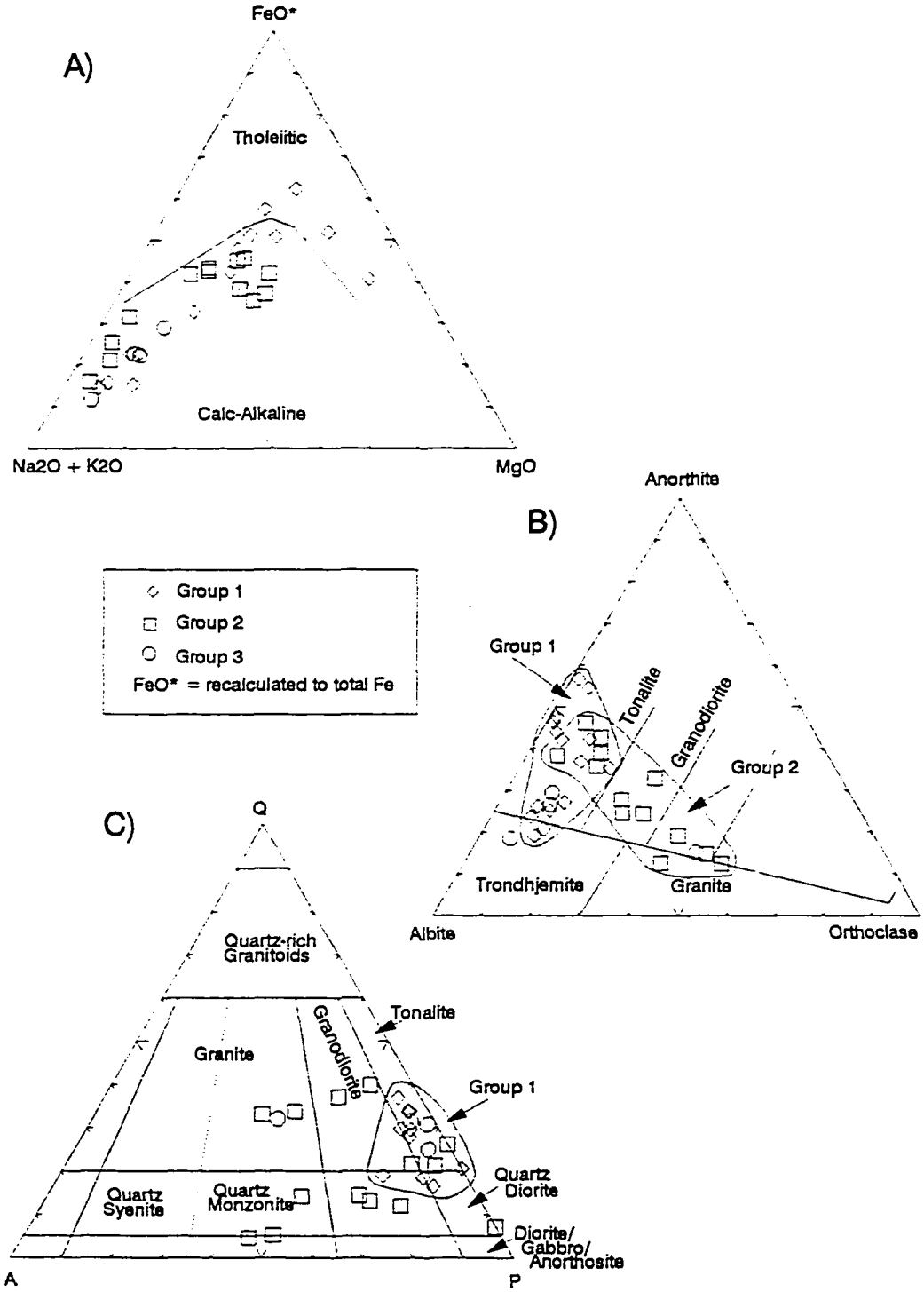


Figure 4. Compositional classification of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - AFM diagram (Irvine & Baragar, 1971). FeO* = total Fe calculated as FeO. **(B)** - Anorthite-Albite-Orthoclase diagram (fields from Barker, 1979). **(C)** - CIPW mesonormative quartz-alkali-feldspar-plagioclase diagram from Le Maitre (1989).

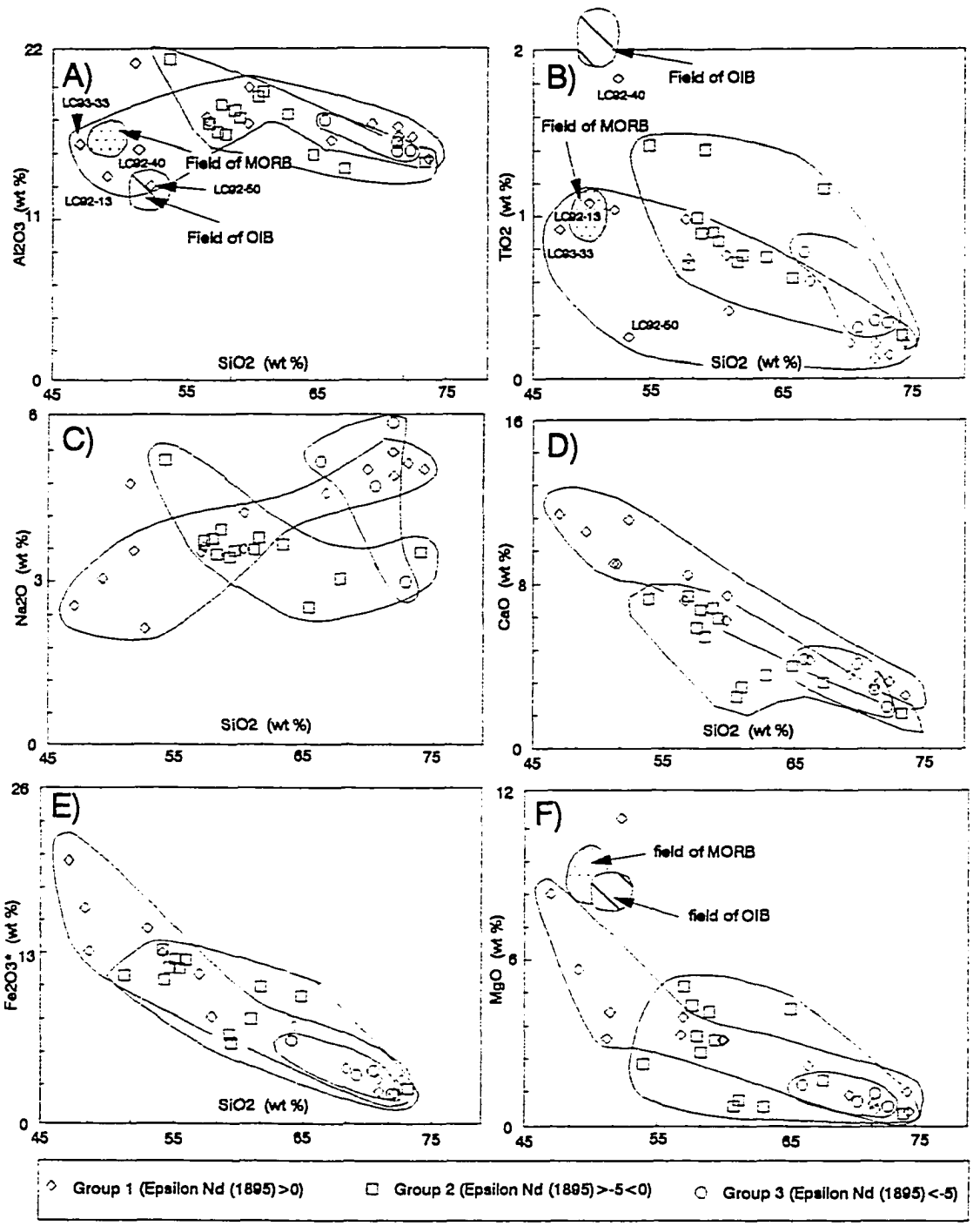


Figure 5. Selected major elements vs. SiO_2 for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Fields of ocean island basalts (OIB) and mid-oceanic ridge basalts (MORB) from Wilson (1989).

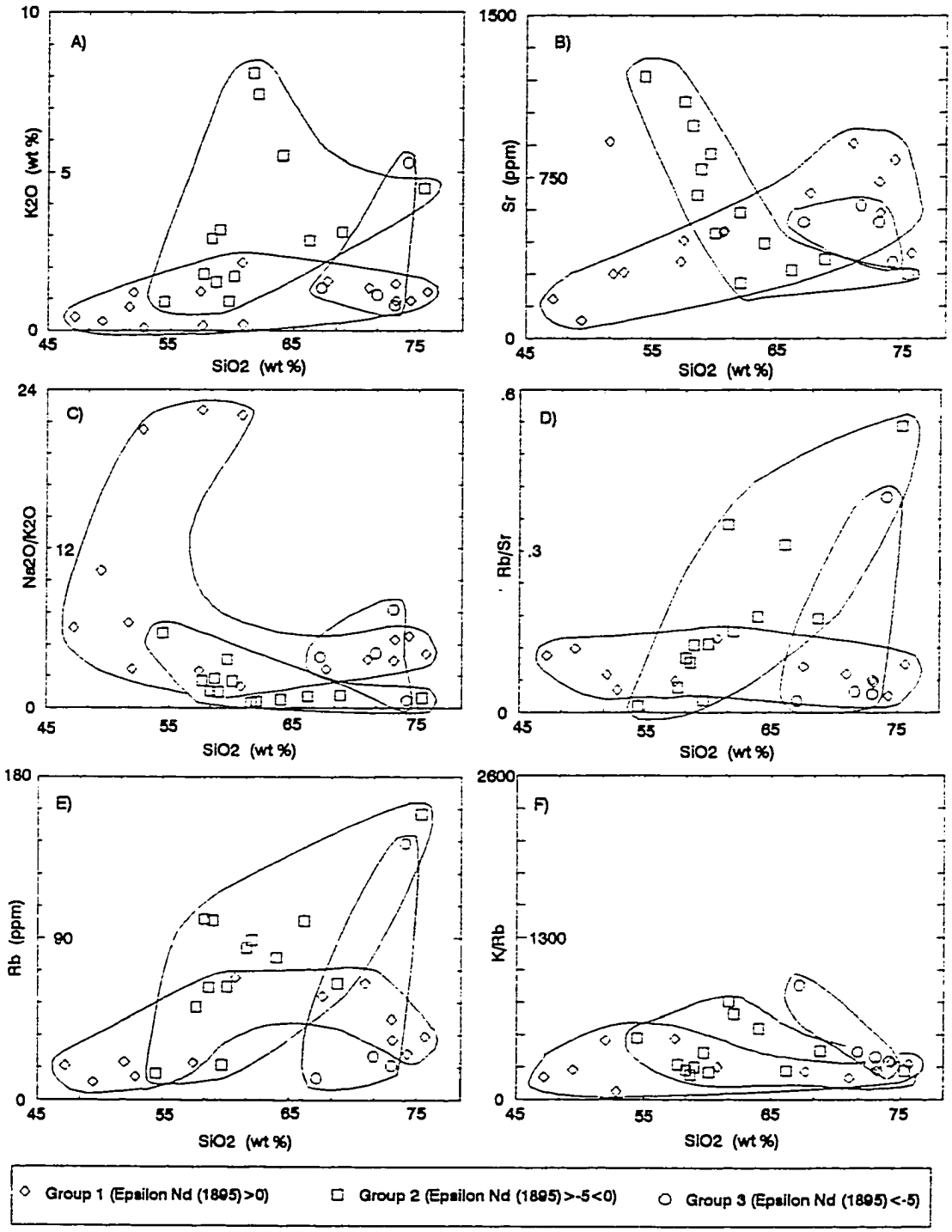


Figure 6. Selected major and trace elements vs. SiO_2 for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador.

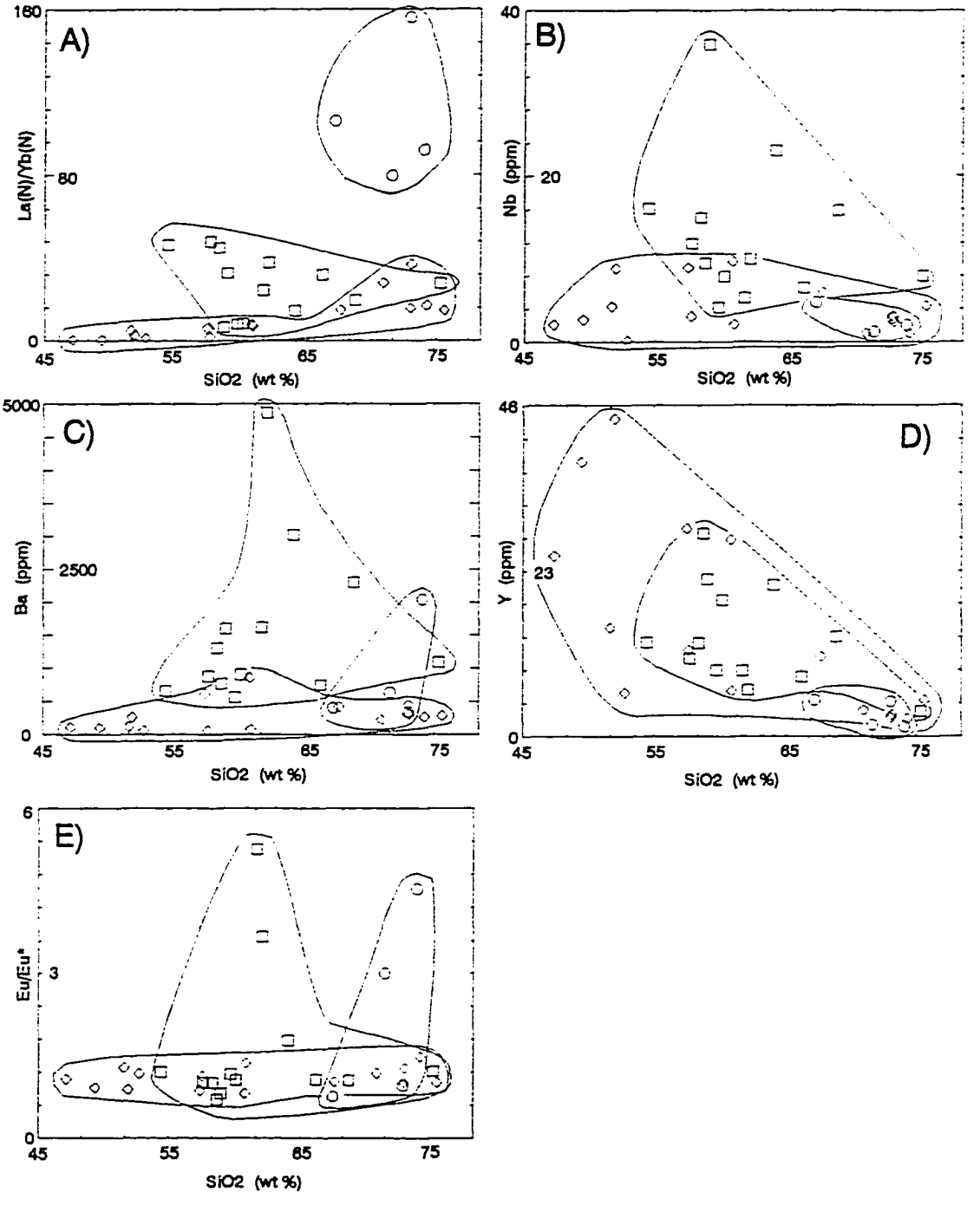


Figure 7. Selected trace and rare earth elements vs. SiO_2 for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador.

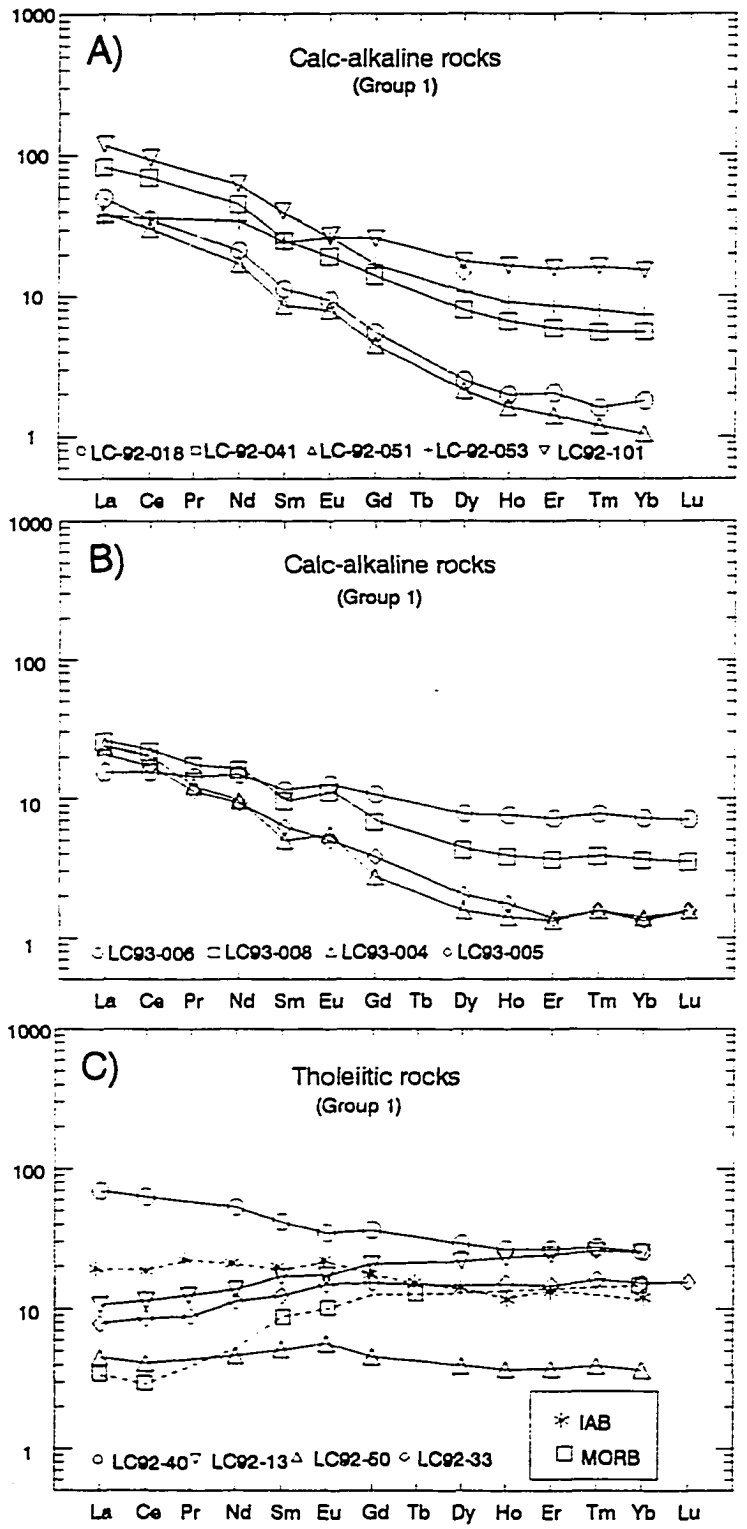


Figure 8. Chondrite normalized REE patterns from Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). **(A)+(B)** - Calc-alkaline rocks in Group 1 ($\varepsilon_{Nd1895} > 0$). **(C)** - Tholeiitic rocks in Group 1 ($\varepsilon_{Nd1895} > 0$). Data from island arc basalt (IAB) and mid-ocean ridge basalt (MORB) from Wilson (1989).

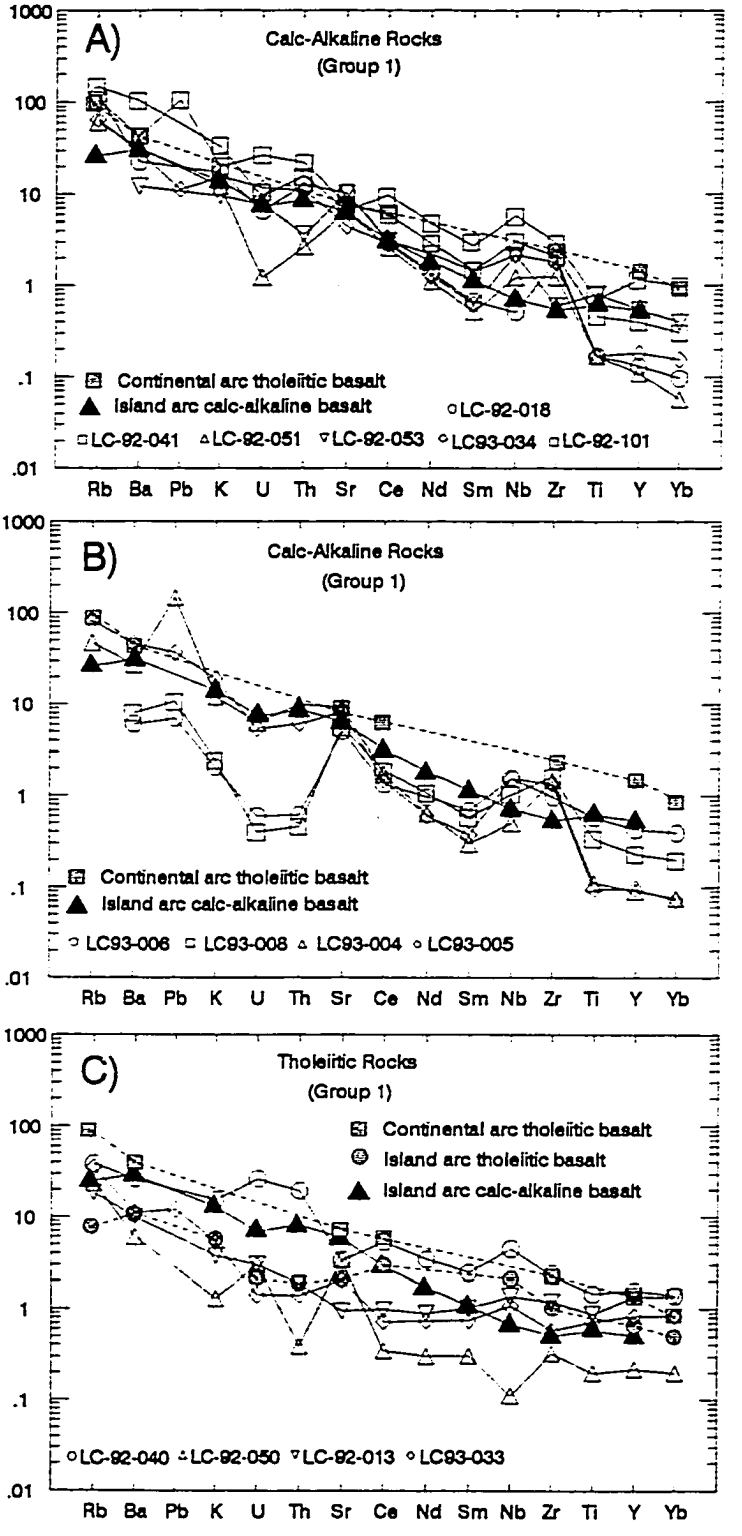


Figure 9. N-MORB normalized trace element data (Spider diagram) for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values and sequence of variables from McCulloch and Gamble (1990). **(A)+(B)** - Calc-alkaline samples in Group 1. **(C)** - Tholeiitic samples in Group 1. Data for continental arc basalt from Ewart (1982), data for island arc tholeiitic basalt from Sun (1980), and data for calc-alkaline island arc basalt from Sun (1980).

(McCulloch and Gamble, 1991). The MORB-normalized values of U and Th are unusually variable among the Group 1 samples, relative to other elements.

3.5.1.2 Geochemistry of Group 2 samples ($\varepsilon Nd_{(1895)}$ = -5 to 0)

All Group 2 samples are calc-alkaline (Fig. 4A), and vary in normative composition from tonalite and granodiorite through granite and quartz monzonite (Figs. 4B and C). SiO_2 ranges from 53 to 74 wt.% and Al_2O_3 decreases with increasing SiO_2 from 14.5 to 21 wt.%. MgO, Fe_2O_3, CaO and TiO_2 all decrease with increasing SiO_2, but the data are more scattered than those from Group 1. Sr and Na_2O both decrease sharply as SiO_2 increases from 53 to 62 wt.%, but are constant above 62 wt.% SiO_2. K_2O and Rb show a general increase with increasing SiO_2, but the data are scattered. Rb/Sr correlates positively with increasing SiO_2, and Na_2O/K_2O and K/Rb decrease with increasing SiO_2. Ba concentrations are high (570-4800 ppm), but uncorrelated with SiO_2. Nb and Y concentrations are also scattered, but generally decrease with increasing SiO_2 (Figs. 5, 6 and 7).

REEs are moderately to strongly fractionated among the Group 2 samples ($La_{(N)}/Yb_{(N)}$ = 8-50; Fig. 7A), and several samples have large positive Eu anomalies (Fig. 7E and 10A). In Fig. 10, subdivision of Group 2 samples is based on the degree of REE fractionation coupled with the presence or absence of a positive Eu anomaly. The overall REE concentrations of Group 2 samples are higher than for Group 1. This is particularly true for LREE, which range from ca. 40 to 350 times chondrite for Group 2 (Fig. 10).

On N-MORB normalized spider diagrams (Figs. 11A-C), Group 2 samples show a general decrease in element abundances from left to right. The samples have greater than MORB concentrations in all elements except Ti, Y and Yb, which range between 0.1 and 1 times MORB-values. The rocks are strongly enriched in Rb, Ba, Pb and K (Ba

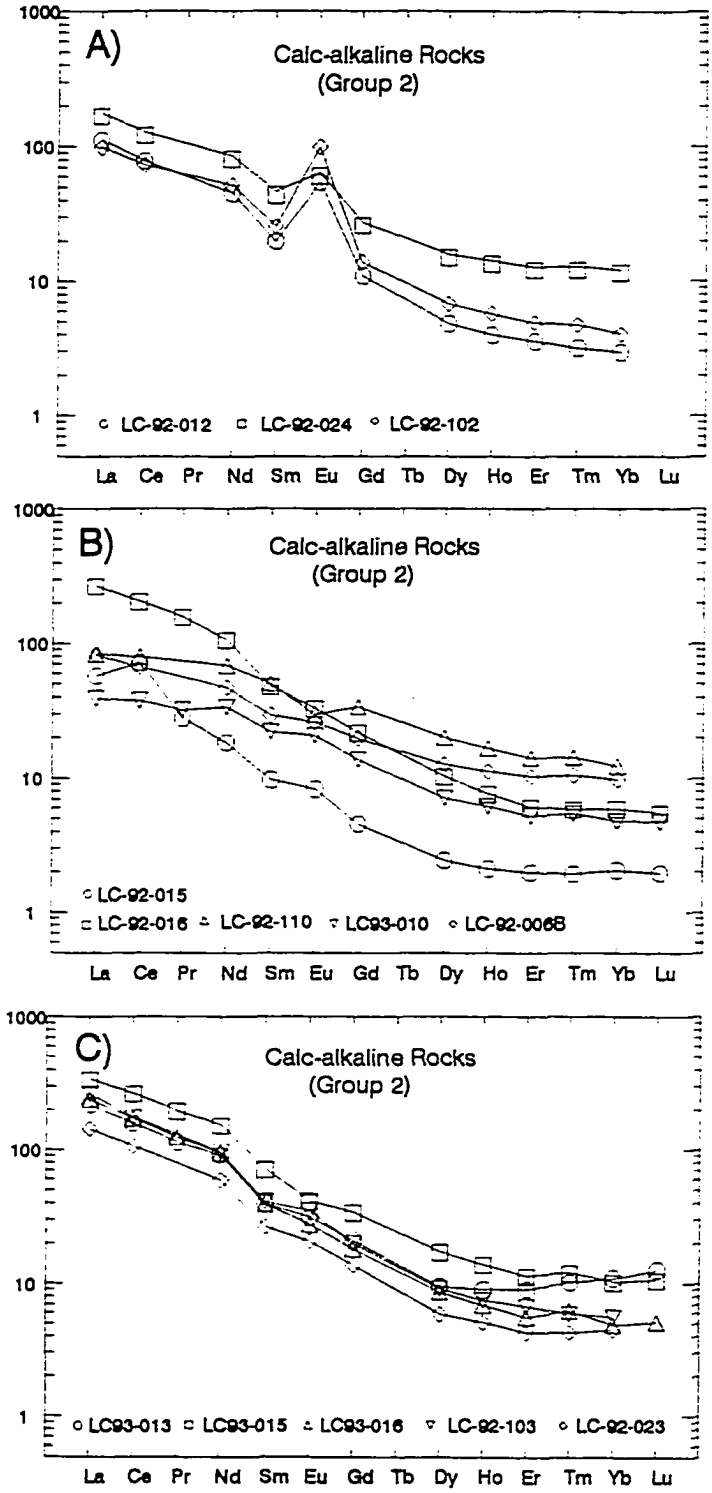


Figure 10. Chondrite normalized REE patterns for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values from Evensen (1978). All data from Group 2 samples ($-5 < \varepsilon_{Nd1896} < 0$). Samples distributed in (A), (B) and (C) according to shape of REE patterns.

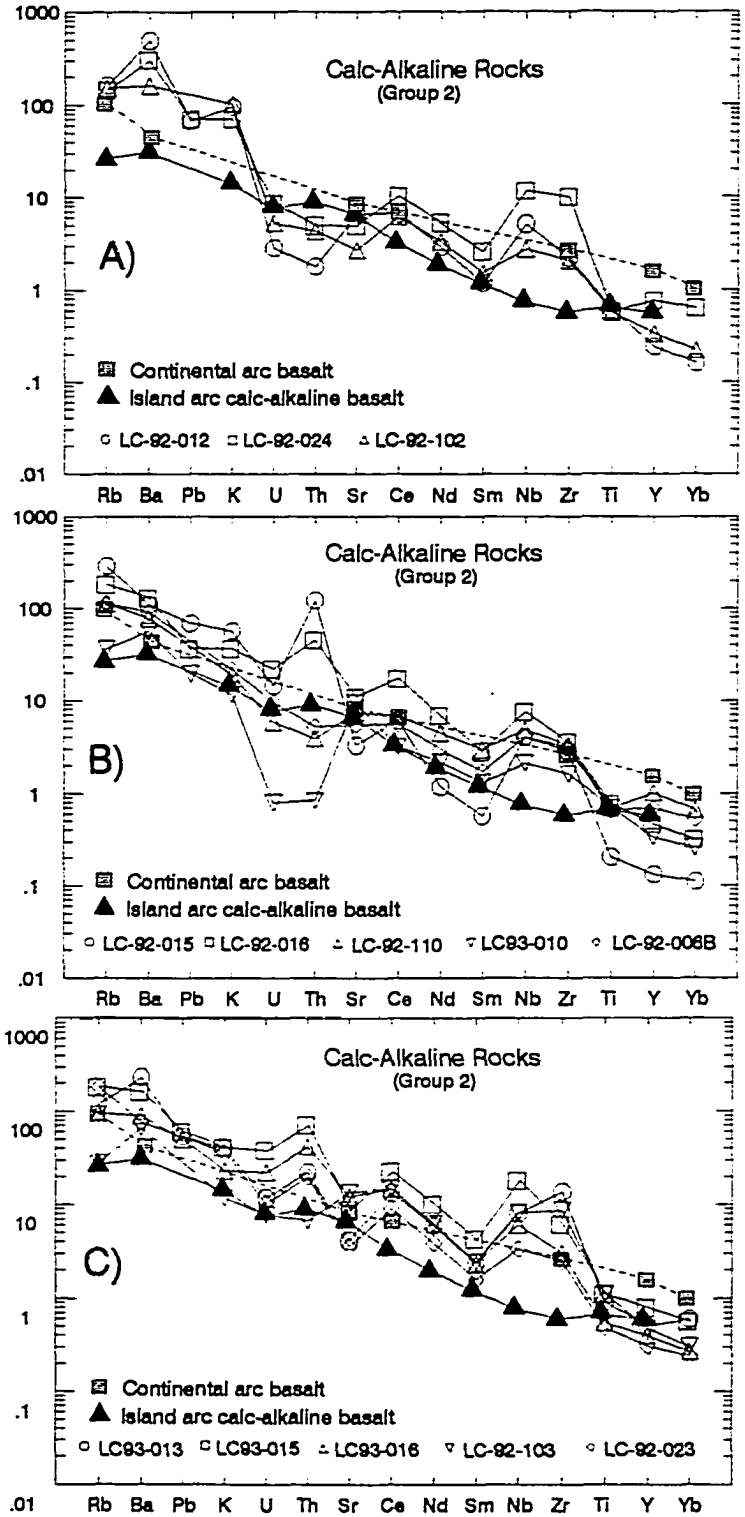


Figure 11. N-MORB normalized trace element data (Spider diagram) for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Normalizing values and sequence of variables from McCulloch and Gamble (1990). All data from Group 2 samples ($-5<\varepsilon_{Nd(1895)}<0$). Distribution of samples in (A), (B) and (C) based on shapes of REE patterns (see Figure 9). Data for continental arc basalt from Ewart (1982) and data for island arc calc-alkaline island arc basalt from Sun (1980).

ranges from 60 to 600 times MORB-values). U and Th concentrations are extremely variable, ranging from ca. 1 to 100 times MORB-values, and the samples are typically enriched in Nb and Zr, which range from 2 to 10 times MORB-values.

3.5.1.3 Geochemistry of Group 3 samples ($\varepsilon Nd_{(1895)} < -5$)

Group 3 samples are tonalites, trondhjemites and granites in CIPW-normative diagrams (Figs. 4B and C). They are calc-alkaline (Fig. 4A), and range from 65.1 to 70.6 wt.% SiO_2. Al_2O_3 ranges from 14.9 to 17.1 wt.%. Fe_2O_3, MgO, TiO_2, CaO, Sr and K/Rb all decrease with increasing SiO_2 (Figs. 5, 6 and 7). The samples have low Nb and Y values that are similar to those in Group 1 (Fig. 7). With the exception of LC92-109, Group 3 samples have low K_2O, Rb, Ba, Rb/Sr, Nb and Y, with values similar to those for felsic samples in Group 1.

Group 3 samples have strongly fractionated REEs with $La_{(N)}/Yb_{(N)}$ between ca. 80 and 160 (Fig. 7A), and several have strong positive Eu anomalies (Figs. 7E and 12A). LREE concentrations range between 50-300 times chondrite, whereas HREEs range between 0.6-2.5 times chondrite (Fig. 12A).

In the N-MORB-normalized spider diagram (Fig. 12B), Group 3 samples show strong enrichments in Rb, Ba, Pb and K with the same overall decrease in normalized element abundances from left to right, as observed for Groups 1 and 2. U and Th concentrations are extremely variable. Y and Yb are strongly depleted relative to N-MORB.

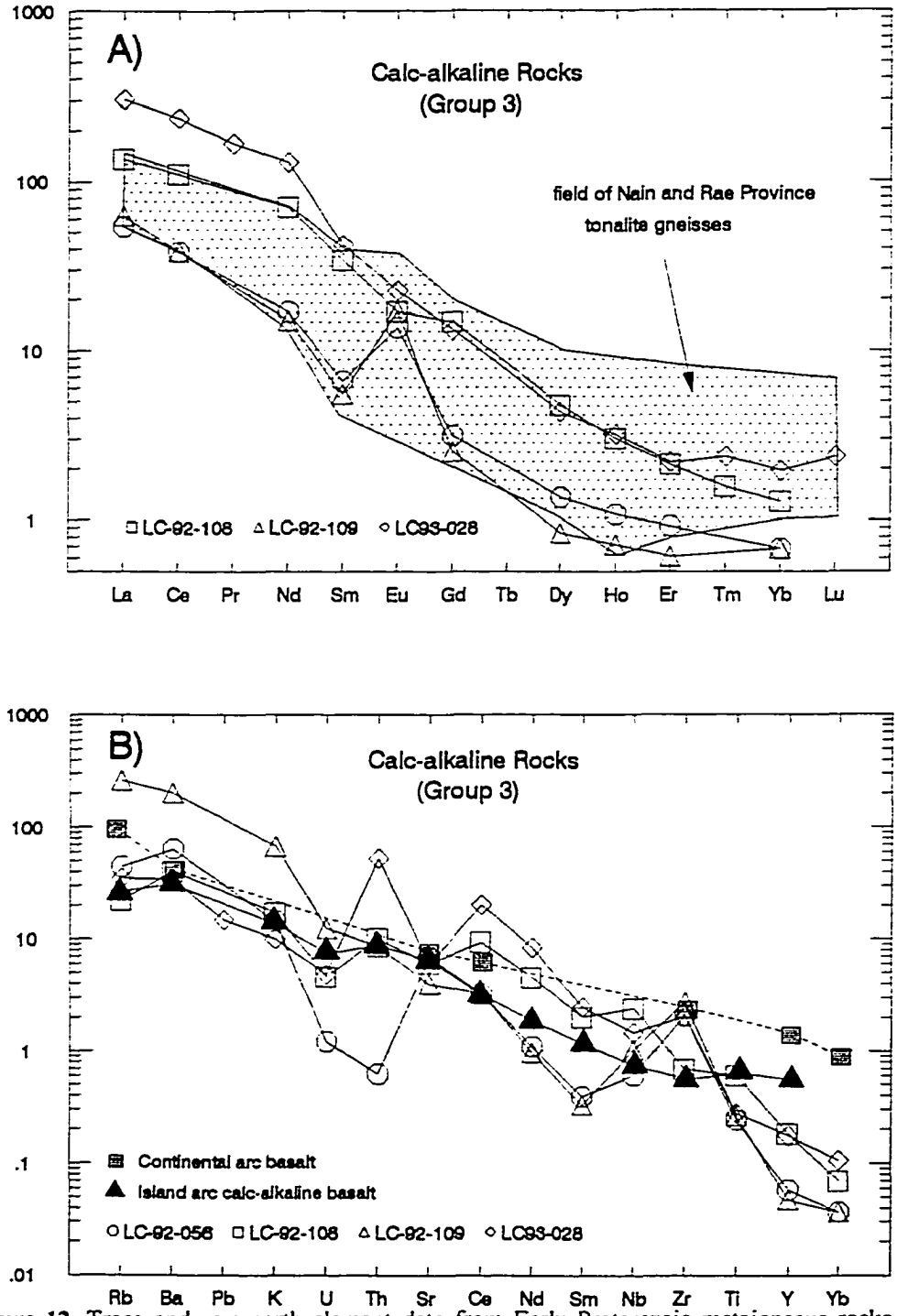


Figure 12. Trace and rare earth element data from Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - Chondrite normalized REE patterns from Group 3 samples ($\varepsilon_{Nd(1895)} < -5$). Normalizing values from Evensen (1978). Shaded field represent compositional variation among Nain and Rae Province Archean gneisses (data from Chapter 2). **(B)** - N-MORB normalized trace element data from Group 3 samples ($\varepsilon_{Nd(1895)} < -5$). Normalizing values and sequence of elements from McCulloch and Gamble (1990). Data for continental arc basalt from Ewart (1982) and data for island arc calc-alkaline island arc basalt from Sun (1980).

3.5.2 Isotope Geochemistry

Neodymium, strontium and lead isotopic data for selected whole rock samples are summarized in Table 3 and Figs. 13-17. Analytical details are given in Appendix A. Stacey-Kramers two-stage evolution curve is shown for reference to the $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ data (Fig. 15; Stacey and Kramers, 1975).

3.5.2.1 Group 1 ($\varepsilon Nd_{(1895)} > 0$)

Calc-alkaline samples in Group 1 have $\varepsilon Nd_{(1895)} > 0$ and $\varepsilon Nd_{(M)}$ values ranging from -5.56 to -20.9, with $^{147}Sm/^{144}Nd$ ratios between 0.1002 and 0.1592 (Figs. 13A and B; Table 3). Tholeiitic samples (gabbro LC92-50 and amphibolites LC92-13 and -33) have $\varepsilon Nd_{(M)}$ values of 3.32 to 14.34, with $^{147}Sm/^{144}Nd$ ratios of 0.2023-0.2430. Collectively, all Group 1 samples form a linear array that plots along a 1.967 Ga reference isochron (Fig. 13B). Depleted mantle source compositions calculated using data from Collerson et al. (1991) and Stern et al. (1994) for Nain and Superior provinces are shown on Figs. 13A and B for comparison.

Group 1 has distinctly unradiogenic $^{87}Sr/^{86}Sr$ and low $^{87}Rb/^{86}Sr$ (Figs. 14A and B) relative to samples in Groups 2 and 3. The samples form a broadly linear array coincident with a reference isochron of 2.14 Ga (Fig. 14B), with an $^{87}Sr/^{86}Sr$ intercept of 0.702184.

There is a large range in $^{206}Pb/^{204}Pb$ (15.644-20.787)and $^{207}Pb/^{204}Pb$ (15.135-15.736) values among Group 1 (Fig. 15A). $^{206}Pb/^{204}Pb$ shows an overall positive correlation with $^{208}Pb/^{204}Pb$ (Fig. 16A), but the data define a very broad array.

3.5.2.2 Group 2 ($-5 < \varepsilon Nd_{(1895)} < 0$)

$\varepsilon Nd_{(1895)}$ for Group 2 ranges between 0 and <-5, $\varepsilon Nd_{(M)}$ ranges between -14.96 and -33.53 with $^{147}Sm/^{144}Nd$ ratios between 0.0787 and 0.1404. Collectively, $\varepsilon Nd_{(M)}$ vs.

TABLE 3 - Isotopic analyses of Early Proterozoic metaigneous rocks, Northern Labrador

	Rb (ppm)	Sr (ppm)	87Rb/ 86Sr	87Sr/ 86Sr(M)	Sm (ppm)	Nd (ppm)	147Sm/ 144Nd	143Nd/ 144Nd(M)	143/144Nd (+/-2 sigma)	ENd (M)	ENd (1895 Ma)	T(DM) (Ga)	206Pb/ 204Pb	207Pb/ 204Pb	208Pb/ 204Pb
Group 1 (Epsilon Nd(1895)>0)															
LC92-018	49.8	758	0.1900	0.70714	1.68	10.0	0.1023	0.511620	10	-19.86	3.31	1.95	15.881	15.135	36.216
LC92-041	29.7	514	0.1668	0.70857	3.36	18.5	0.1101	0.511628	10	-19.70	1.55	2.08	18.282	15.451	36.714
LC92-051	19.0	495	0.1110	0.70550	1.31	7.9	0.1002	0.511618	18	-19.90	3.78	1.91	15.644	15.143	35.707
LC92-053	4.6	799	0.0168	0.70266	3.69	16.0	0.1392	0.512084	12	-10.81	3.32	1.96	16.961	15.338	36.408
LC92-101					5.82	29.7	0.1193	0.511656	5	-19.16	-0.59	2.24	16.388	15.399	36.162
LC93-004					0.85	5.0	0.1035	0.511566	7	-20.91	1.96	2.04			
LC93-005	35.2	685	0.1485	0.70623	0.92	4.3	0.1284	0.511900	19	-14.40	2.36	2.04	16.547	15.286	36.128
LC93-006	1.1	442	0.0074	0.70239	1.76	6.7	0.1592	0.512353	15	-5.56	3.58	1.92	16.984	15.409	36.729
LC93-008					1.49	7.2	0.1242	0.511880	6	-14.79	2.86	1.98	16.832	15.410	36.648
LC93-011	14.0	341	0.1190	0.70668	5.79	28.2	0.1241	0.511747	20	-17.38	0.28	2.21			
LC93-034	23.0	350	0.1896	0.70825	1.55	8.8	0.1067	0.511681	9	-18.67	3.41	1.94	19.603	15.663	36.888
LC92-013th					2.58	6.4	0.2430	0.513373	1790	14.34	3.00	-13.41	20.787	15.736	38.163
LC93-033th	5.2				1.89	4.8	0.2370	0.513357	9	14.03	4.17	-10.65	17.072	15.353	36.977
LC92-040th		272		0.70653	6.31	24.6	0.1552	0.512320	5	-6.20	4.00	1.88	19.536	15.552	37.501
LC92-050th					0.68	2.0	0.2023	0.512809	9	3.34	1.97	4.15	18.107	15.453	36.846

"th" = tholeiites; "M" = measured.
Analytical details are given in Appendix 1.

TABLE 3 - (continued) Isotopic analyses of Early Proterozoic metaigneous rocks, Northern Labrador

	Rb (ppm)	Sr (ppm)	87Rb/ 86Sr	87Sr/ 86Sr(M)	Sm (ppm)	Nd (ppm)	147Sm/ 144Nd	143Nd/ 144Nd(M)	143/144Nd (+/-2 sigma)	ENd (M)	ENd (1895 Ma)	T(DM) (Ga)	206Pb/ 204Pb	207Pb/ 204Pb	208Pb/ 204Pb
Group 2 (-5 < Epsilon Nd(1895) < 0)															
LC92-006B					5.02	24.9	0.1219	0.511661	6	-19.06	-1.12	2.30	16.734	15.469	35.759
LC92-012	73.1	502	0.4213	0.71592	3.06	21.4	0.0864	0.511010	11	-31.76	-5.11	2.44	14.659	14.967	34.762
LC92-015		290		0.74318	1.43	9.4	0.0923	0.511126	10	-29.49	-3.92	2.42			
LC92-016		923		0.71141	7.08	54.4	0.0787	0.510919	6	-33.53	-4.63	2.41			
LC92-023		212		0.73275	3.90	26.3	0.0898	0.511173	9	-28.58	-2.37	2.30	16.270	15.389	35.611
LC92-024	49.3	270	0.5281	0.71836	6.58	38.3	0.1040	0.511405	8	-24.05	-1.52	2.28	15.716	15.204	35.082
LC92-102	73.1	201	1.0527	0.73271	3.78	23.2	0.0986	0.511311	6	-25.89	-2.04	2.30	15.282	15.086	34.938
LC92-103	3.6	1471	0.0071	0.70416	5.98	44.7	0.0809	0.511087	10	-30.26	-2.10	2.25	15.799	15.154	35.270
LC92-110					8.33	35.9	0.1404	0.511871	7	-14.96	-1.45	2.45	16.595	15.359	35.701
LC93-010	11.4	865	0.0379	0.70575	3.18	15.1	0.1270	0.511712	12	-18.06	-1.12	2.35			
LC93-013	56.7	345	0.4752	0.71717	5.99	41.1	0.0880	0.511221	10	-27.64	-1.22	2.21	15.730	15.227	35.641
LC93-015					9.90	67.2	0.0892	0.511111	9	-29.79	-3.44	2.37	17.093	15.252	37.617
Group 3 (Epsilon Nd(1895)<-5)															
LC92-056	16.6	515	0.0929	0.70690	1.01	7.7	0.0794	0.510617	12	-39.42	-10.93	2.77	14.523	14.767	35.974
LC92-108	9.0	474	0.0550	0.70798	5.07	32.5	0.0943	0.511019	17	-31.58	-6.71	2.60	15.440	15.111	35.438
LC92-109					0.87	7.2	0.0728	0.510779	11	-36.26	-6.16	2.46	16.638	15.429	34.875
LC93-028					6.22	58.3	0.0645	0.510349	11	-44.65	-12.56	2.77			

"M" = measured
Analytical details given in Appendix 1.

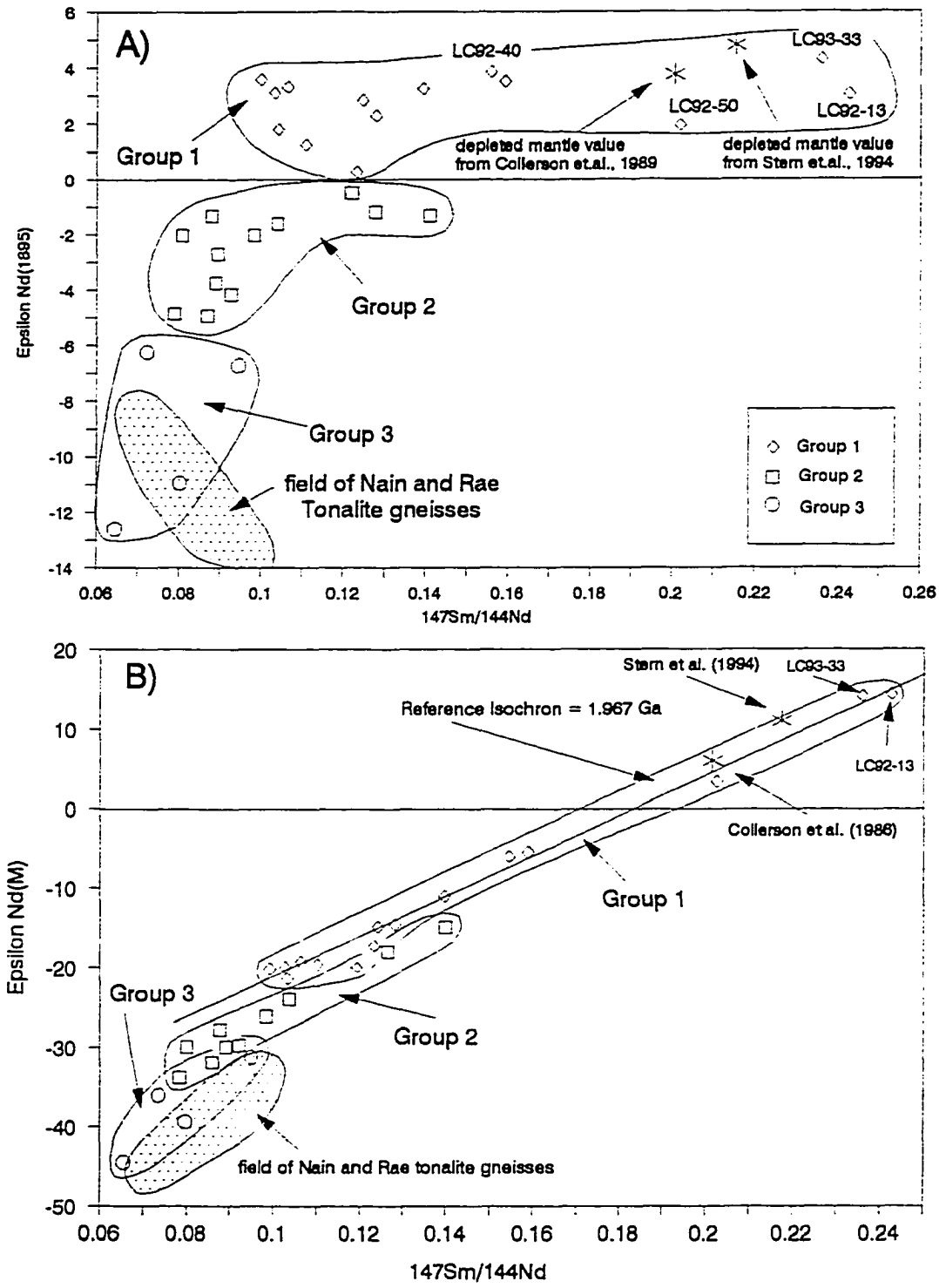


Figure 13. Sm/Nd isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen Northern Labrador. **(A)** - $\varepsilon_{Nd\ (1895)}$ vs. $^{147}Sm/^{144}Nd$ diagram for samples from Group 1, 2 and 3. Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). **(B)** - $\varepsilon_{Nd\ (M)}$ vs. $^{147}Sm/^{144}Nd$ for samples from Group 1, 2 and 3. Stippled field includes data from Nain and Rae Province Archean gneisses (Chapter 2). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994).

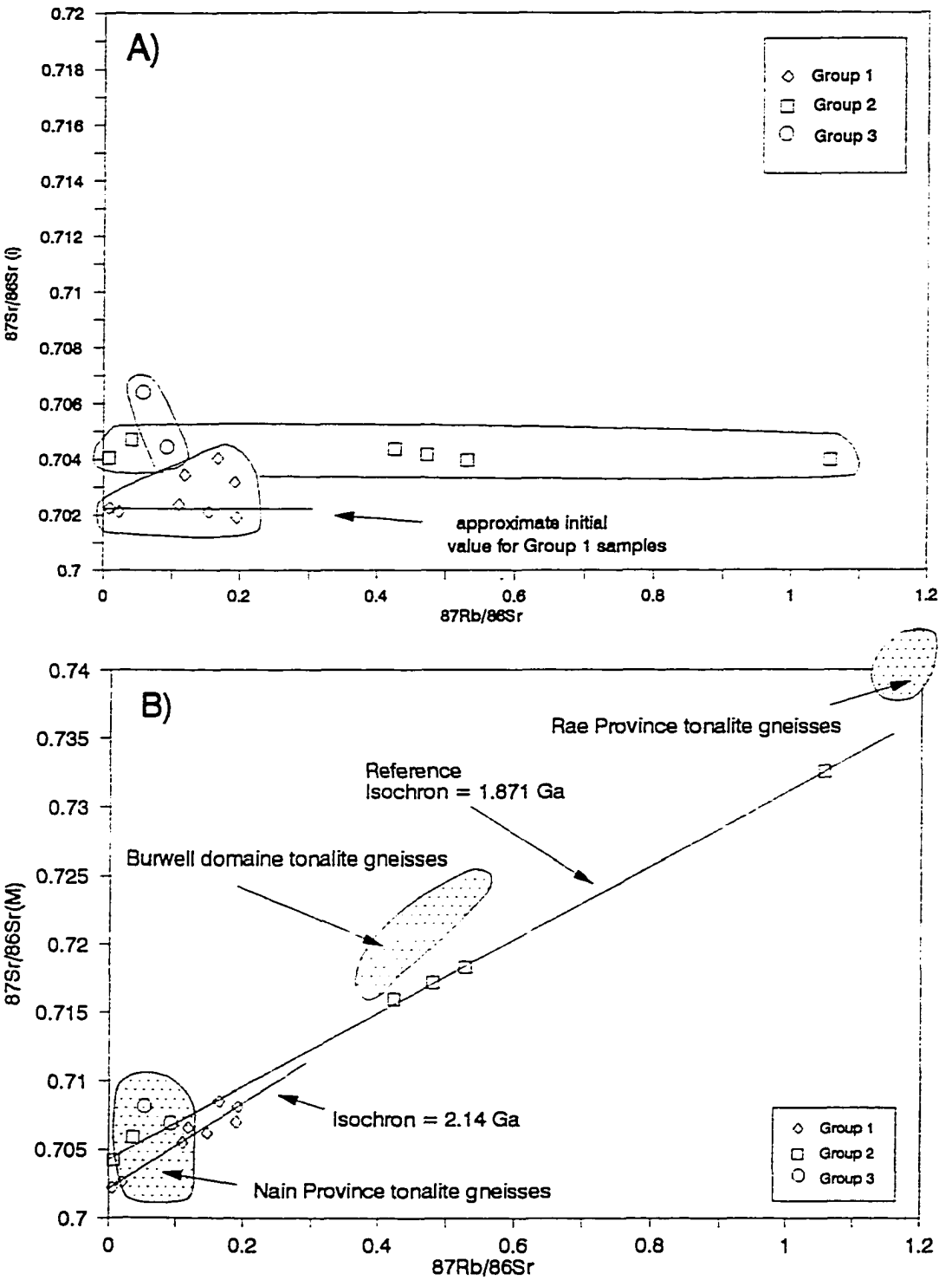


Figure 14. Rb/Sr and Sm/Nd isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - $^{87}Rb/^{86}Sr$ vs. $^{87}Sr/^{86}Sr(I)$ isotope correlation diagram for samples from Group 1, 2 and 3. **(B)** - Rb/Sr isochron diagram for samples from Group 1, 2 and 3. Fields of Nain and Rae Province and Burwell Domain tonalites based on data from Chapter 2. Isochrons at 1.87 Ga and 2.14 Ga are shown for reference.

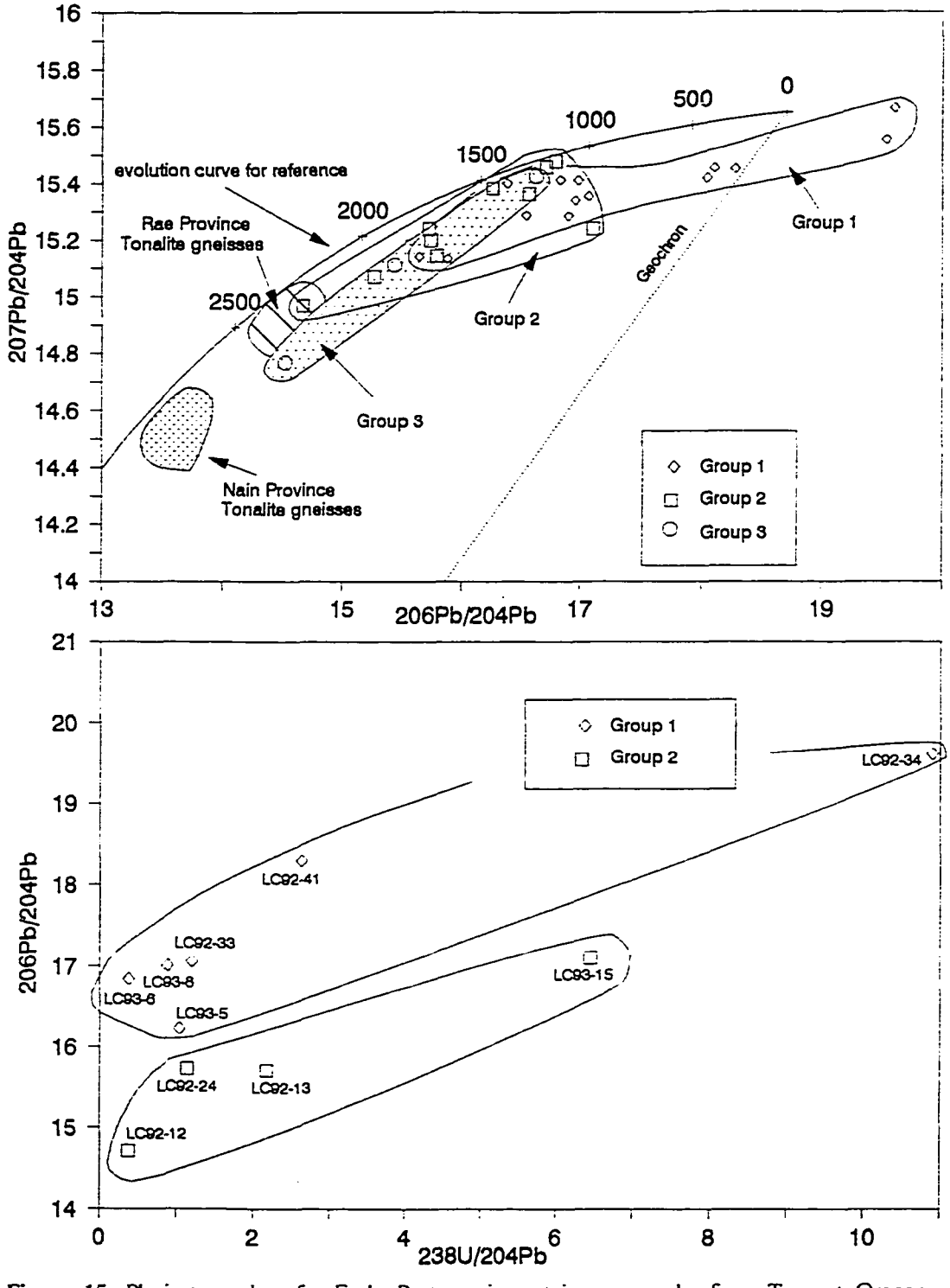


Figure 15. Pb isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ isotope correlation diagram showing the distribution of samples from Group 1, 2 and 3. Also shown are fields of Nain and Rae Province and Burwell Domain tonalites based on data from Chapter 2. Stacey-Kramers two stage evolution curve (Stacey & Kramers, 1975) is shown for reference. **(B)** - $^{238}U/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ isotope correlation diagram showing distribution of Group 1 and 2 samples..

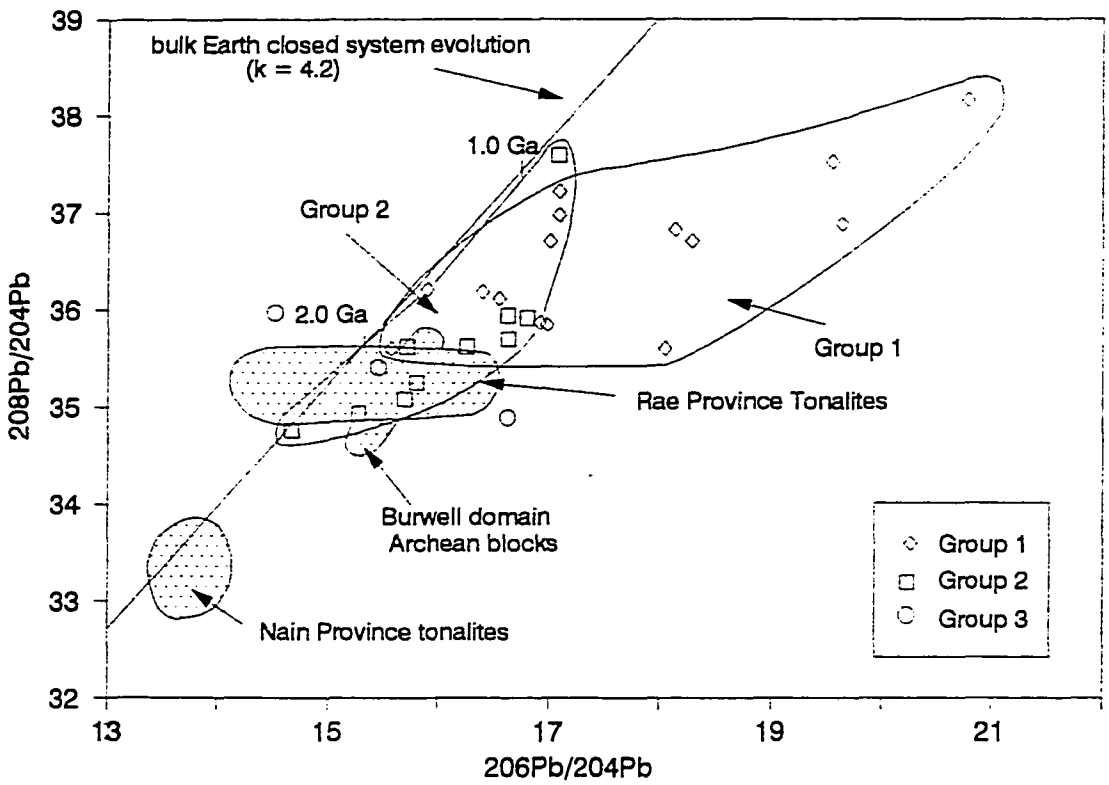


Figure 16. Pb isotope data for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. $^{206}Pb/^{204}Pb$ vs. $^{208}Pb/^{204}Pb$ isotope correlation diagram showing the distribution of samples from Group 1, 2 and 3. Also shown are fields of Nain and Rae Province and Burwell Domain Archean tonalites based on data from Chapter 2.

$^{147}Sm/^{144}Nd$ co-vary with a positive slope (Fig. 13B), at lower $\varepsilon Nd_{(M)}$ values than among Group 1 samples.

Relative to Group 1, $^{87}Sr/^{86}Sr$ ratios are more radiogenic (Figs. 14A and B), and most samples have higher Rb/Sr. In Fig. 14B, Group 2 samples form a tight linear array over a large range in $^{87}Rb/^{86}Sr$ values, which produces a good fit along a 1.87 Ga reference isochron, with an $^{87}Sr/^{86}Sr$ intercept of 0.70396.

The Pb isotopic ratios are generally less radiogenic than those of Group 1, although there is significant overlap between their fields on $^{207}Pb/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ diagrams (Figs. 15A and 16A).

3.5.2.3 Group 3 ($\varepsilon Nd_{(1895)} < -5$)

$\varepsilon Nd_{(1895)}$ values for Group 3 samples are <-5 and $\varepsilon Nd_{(M)}$ values range between -28.58 and -39.42 with $^{147}Sm/^{144}Nd$ ratios between 0.0645 and 0.0943 (Figs. 13A and B). In Fig. 13B, the samples overlap with the field defined by Archean Nain and Rae Province tonalite gneisses (data from Chapter 2). In Fig. 14B, $^{87}Rb/^{86}Sr$ and $^{87}Sr/^{86}Sr$ values overlap with those from the Nain Province Archean gneisses, but are distinct from those of the Rae Province and Burwell domain rocks. Rb/Sr for Group 3 rocks are similar to those of Group 1, but at higher $^{87}Sr/^{86}Sr$ ratios (Fig. 14B).

On a $^{206}Pb/^{204}Pb$ vs. $^{207}Pb/^{204}Pb$ diagram (Fig. 15A), Group 3 samples overlap with the field of Group 2 samples. $^{208}Pb/^{204}Pb$ vs. $^{206}Pb/^{204}Pb$ for Group 3 samples produce a negative slope.

3.6 DISCUSSION

3.6.1 Geochemical Effects of Granulite Facies Metamorphism

Before discussing the geochemical and isotopic characteristics of the Early Proterozoic intrusive rocks in the northern TO, the geochemical effects of high-grade metamorphism must be identifed. Within the northern TO all lithologies are metamorphosed and contain upper amphibolite- to granulite facies peak-metamorphic mineral assemblages. Subsequently, many rocks were retrogressed to amphibolite facies. It is therefore not likely that the present geochemical composition of the rocks is identical to that of their protoliths. Below we evaluate the geochemical effects of the metamorphism experienced by these rocks.

All the Early Proterozoic metaigneous rocks in this study are enriched in Rb, Ba, Pb and K relative to MORB (Figs. 9, 11, 12), however, relative to calc-alkaline island arc basalts (IAB), most of the rocks are not significantly enriched or depleted in these elements. Assuming that the calc-alkaline rocks in Groups 1, 2 and 3 represent arc magmas (discussed below), most of them do not reflect the extreme LILE-depletion commonly observed among granulite facies rocks (e.g., Heier, 1965, 1973, Touret, 1996). K/Rb ratios are considered to be a sensitive indicator for LILE-depletion associated with granulite facies metamorphism (Rudnick et al., 1985; Touret, 1996). In Fig. 17, the rocks from this study do not follow the trends observed among LILE-depleted granulites (Rudnick et al., 1985), and instead plot within the main trend of unmetamorphosed igneous rocks developed by Shaw (1968). This further indicates that the rocks were not significantly depleted in LILEs during granulite facies metamorphism.

It is possible that the rocks were depleted, then subsequently re-enriched in the LILEs during later retrogression (e.g., Rudnick et al., 1985), however, the Sr isotope data do not support this. The good fit of datapoints for Group 1 and 2 samples along Rb-

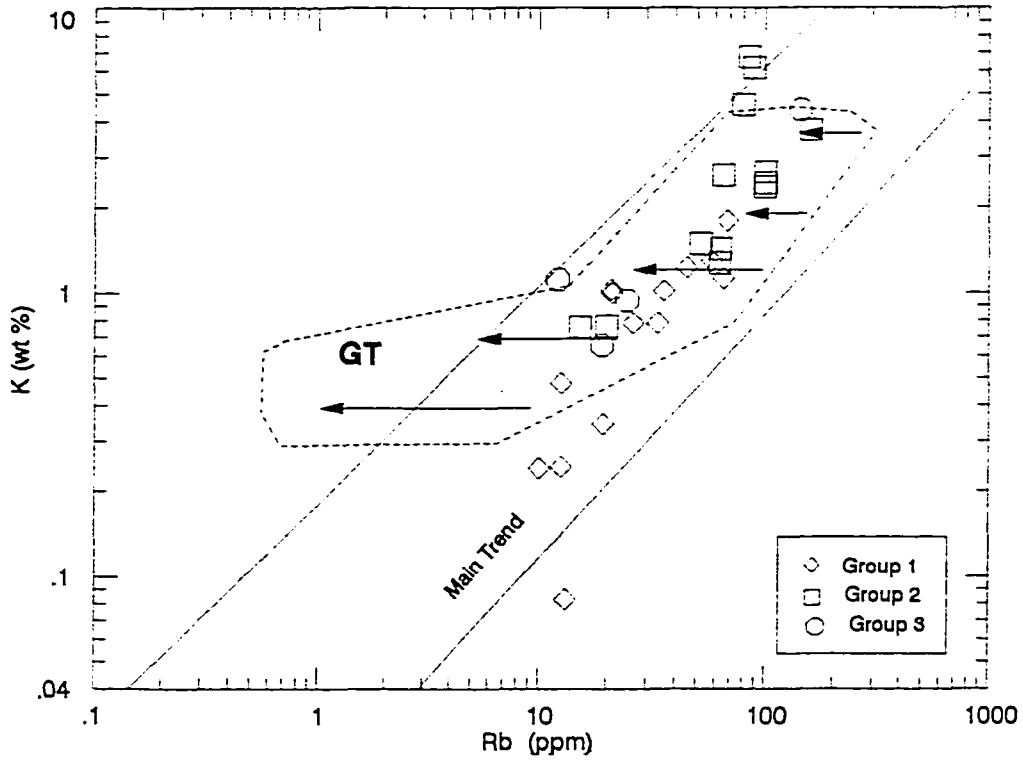


Figure 17. Plot of Rb (ppm) versus K (wt.%) for Early Proterozoic metaigneous rocks from Torngat Orogen. The diagram also shows the "Granulite Trend" (GT) from Rudnick et al. (1985) and the "Main Trend" of continental igneous rocks (granites through basalts) from Shaw (1968). Horizontal arrows represent the amount and direction of Rb depletion caused by granulite facies metamorphism (from Rudnick et al., 1985).

Sr reference isochrons (2.14 Ga and 1.87 Ga, respectively; Fig. 14A), suggests that the Rb-Sr systems for these samples were not significantly disturbed after crystallization of the rocks.

In contrast to this observation, the geochemically primitive, tholeiitic rocks in this study (LC92-13, -50 and LC93-33) have MORB-like concentrations for most trace elements, but the high Rb, Ba and Pb values relative to N-MORB deviate from the otherwise depleted character of the rock, suggesting that these elements were introduced during secondary events, however it is not impossible that the mantle source of the amphibolites was metasomatized and enriched in LILEs.

On the basis of the comparison with unmetamorphosed IAB (Sun, 1980) and MORB (McCulloch and Gamble, 1990), we suggest that the geochemical effects of metamorphism were in general relatively minor, and did not significantly alter the chemistry of their protoliths, however, a few of the samples (particularly the tholeiites) show secondary enrichment in some LILEs, particularly Rb and Ba.

3.6.2 Origin of Group 1 rocks

3.6.2.1 Tholeiites

Amphibolites

The tholeiitic amphibolites in this study are similar to N-MORB in most aspects of their geochemistry, with the exception of the LILE enrichment. This conclusion is based on a number of geochemical features such as major element concentrations (Figs. 5A, B and F), Zr/Nb ratios (20-33, not shown) and REE patterns (ca. 10 times chondrite with $La_{(N)}/Yb_{(N)} < 1$), all of which correspond to values for N-MORB (Zr/Nb = 30; Pearce, 1983), but differ from those of ocean island basalts (Zr/Nb = 10), and of continental arc basalts (Zr/Nb = ca. 13; Wilson, 1989). In addition, the amphibolites

have HFSE- and REE patterns which are similar to N-MORB, but which differ from tholeiitic island arc- and continental arc basalts (Fig. 9C).

The amphibolites show MORB-like concentrations of the elements Ce through Yb in the spider diagram (Fig. 9C). This is coupled with $\varepsilon Nd_{(1895)}$ values which correspond to a depleted mantle source (Figs. 13A and B), suggesting that the REE and HFSE character of the amphibolites is an igneous feature not substantially modified by later events.

The N-MORB-like chemistry of the amphibolites and their field occurrence as tectonically bound sheets suggest that they may have originated as allochthonous oceanic crust or oceanic plateau (Condie, 1997), which was tectonically intercalated with its host rocks (further discussed below).

Cumulate Gabbro

The cumulate gabbro (LC92-50) has low $\varepsilon Nd_{(1895)}$ (= 1.97), relative to its presumed depleted mantle source (cf., depleted mantle values of 3.92 and 4.4 calculated with data from Collerson et al. (1991) and Stern et al. (1994), respectively; Fig.13A). This suggests that the gabbro was contaminated by material with an $\varepsilon Nd_{(1895)}$ value lower than that of depleted mantle, most likely represented by Nain Province basement gneisses into which the gabbro was emplaced. The gabbro was subsequently cut by a tonalite (LC92-51; $\varepsilon Nd_{(1895)}$ = 3.78). The high MgO and Sc concentrations (10.5 wt.% and 46 ppm, respectively) and overall low REE- and HFSE concentrations indicate that the gabbro represents relatively early stages of fractional crystallization from a mantle derived basaltic melt. However, the low TiO_2 and relatively low Cr and Ni concentrations (190 and 80 ppm, respectively), indicate that it is not a primary mantle melt (Wilson, 1989). The high Rb, Ba and U values relative to N-MORB contrast with the otherwise depleted character of the rock, suggesting that these elements were

introduced during secondary events as discussed above. Sr-enrichment (relative to adjacent elements) reflects accumulation of plagioclase.

Gabbro diorite: a proxy for the basaltic mantle-derived source?

Group 1 includes a large, relatively undeformed intrusive gabbro-diorite body (LC92-40), which is intruded by ca. 1895 Ma tonalites (Scott, 1995a). LC92 -40 has an $\varepsilon Nd_{(1895)}$ value of 3.99, similar to that of depleted mantle discussed above and high REE-concentrations (ca. 30-80 times chondrite), with a minor LREE-enrichment. The enrichment of LC92-40 in LILEs and LREEs relative to MORB (Fig. 9C), are characteristic of continental arc basalts (CAB) (e.g., Hawkesworth and Powell, 1980; Arculus and Powell, 1986; Hildreth and Moorbath, 1988). Low MgO, Ni and Cr indicate that the rock is not a primary mantle melt (Wilson, 1989), but may be related to it by minor fractional crystallization and separation of olivine +/- orthopyroxene. Otherwise, the geochemical character of LC92-40 provides a best estimate for the primary melt from which other Group 1 samples may have evolved.

3.6.2.2 Calc-alkaline rocks (Group 1)

Calc-alkaline members of Group 1 show a systematic covariation of most major elements with SiO_2 suggesting that they are linked by similar magma evolution trends. The negative correlations between TiO_2, CaO Fe_2O_3 and MgO with SiO_2 suggest fractionation of olivine, pyroxenes, calcic plagioclase +/- titanite, +/- hornblende. The systematic decrease in MORB-normalized trace element concentrations (from left to right in Figs. 9A and B), suggests that these element enrichments are controlled by metasomatism of the mantle source by subduction zone fluids and by partial melts (McCulloch and Gamble, 1991), thus supporting an origin as arc magmas.

Generally high initial $^{143}Nd/^{144}Nd$ ratios for Group 1 calc-alkaline rocks indicate no or minor contamination by the Archean basement gneisses into which many of them were clearly emplaced. The tight linear array defined in the $^{143}Nd/^{144}Nd$ vs. $^{147}Sm/^{144}Nd$ diagram (Fig. 13B), indicates that the Nd isotope system has remained coupled with Sm/Nd throughout the history of these rocks. Despite a large range in $^{87}Rb/^{86}Sr$, the $^{87}Sr/^{86}Sr_{(i)}$ values for many of the Group 1 samples are close to 0.70207 (Fig. 14A), which is interpreted to approximate the $^{87}Sr/^{86}Sr_{(i)}$ value for the source of Group 1 rocks. Values slightly higher than this reflect minor crustal contamination or secondary mobility of Rb +/- Sr associated with metamorphism. The high $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ values for a given U/Pb ratio (Fig. 15B) relative to Group 2 samples, further support the lack of significant contamination by Archean crust, since all Archean crust in the area has very unradiogenic Pb and therefore any mixing between mantle derived melt and Archean crust would lower the $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ values.

Based on these isotopic constraints, the coherent linear trends seen in Harker plots and the distinctive trace element character observed in the spider diagrams, we suggest that the protoliths for Group 1 calc-alkaline rocks represent continental arc magmas which evolved through fractional crystallization of a primary basaltic magma at mid-crustal levels.

3.6.3 Origin of Group 2 Rocks

Among Group 2 samples, the covariation of major and trace elements with SiO_2 is quite scattered. The geochemical scatter on Harker covariation plots is equally prominent among compatible elements (e.g., TiO_2, CaO and MgO) and incompatible elements (K_2O, Na_2O, Rb, Sr; Figs. 5, 6 and 7), suggesting they are the result of magmatic rather than secondary processes, possibly caused by varying amounts of

assimilation of basement gneisses which were not subsequently homogenized into the melts.

The covariation of Al_2O_3, Na_2O and Sr with SiO_2 shows markedly different slopes relative to those produced by Group 1 samples, indicating differences in the fractional crystallization history. Among Group 1 samples, the covariation of Al_2O_3 with SiO_2 shows a typical convex pattern where an inflection marks the onset of plagioclase fractionation (Fig. 5A). In contrast, Group 2 shows a continuous negative covariation of Al_2O_3 with SiO_2, indicating either early onset of feldspar fractionation or that feldspar fractionation was preceded by that of another aluminous phase, possibly hornblende. Similarly, Group 1 produces a positive slope for Na_2O and Sr vs. SiO_2 whereas in Group 2, the data produce a negative slope (Figs. 5 and 6). This could result from fractionation of hornblende and calcic plagioclase at lower SiO_2, followed by fractionation of sodic plagioclase at higher SiO_2.

The Nd, Sr and Pb isotopic data for Group 2 samples all support involvement of Archean basement gneiss in their generation. Relative to Group 1 samples, Nd isotopic values for Group 2 are significantly lower (Figs. 13A and B), $^{87}Sr/^{86}Sr_{(i)}$ ratios are higher (Fig. 14A), and $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ ratios are lower, supporting contamination of Group 2 magmas by local Archean crust, which has very unradiogenic Pb ratios (Figs. 15A, 15B and 16). In addition, the $^{87}Sr/^{86}Sr_{(i)}$ data for Group 2 are remarkably constant over a large range in Rb/Sr (Fig. 14A), indicating that Sr was isotopically homogenized, either by mixing between the mantle derived magmas and crustal assimilants, or during subsequent regional metamorphism. Group 1 and 2 rocks commonly occur within < 10 km of each other (Fig. 3), so it is unlikely that Group 2 would have been preferentially affected by pervasive regional metamorphic effects not experienced by Group 1 rocks. It is considered more likely that the differences between Groups 1 and 2 stem from variations in igneous processes, for example differences in the

amount of crustal material assimilated. A factor which may contribute to the observed geochemical and isotopic differences is the presence or introduction of variable amounts of H_2O ($\pm CO_2$) to the magmas. Higher contents of volatiles would affect the sequence and identity of crystallizing phases and could facilitate increased assimilation and partial melting of the Archean basement, such as we observe in Group 2 rocks relative to Group 1.

3.6.4 Origin of Group 3 Rocks

Due to the limited number of samples and the small range in SiO_2, Group 3 cannot be discussed in terms of magmatic evolution trends. With the exception of REE patterns and high K and Ti concentrations, the rocks are geochemically similar to felsic members of Group 1. REE patterns for Group 3 samples are strongly fractionated and commonly have large positive Eu anomalies (Fig. 12). These patterns are unlike those from any of the other Early Proterozoic magmatic rocks in the area, but are indistinguishable from Archean basement gneisses in both Nain and Rae Provinces (see Chapter 2).

As discussed previously, Group 3 samples were selected as examples of Early Proterozoic magmatic rocks on the basis of their lithological and structural field characteristics and by the fact that identical rocks in the area yield igneous U-Pb zircon ages of ca. 1910-1869 Ma (Scott, 1995a; Scott and Machado, 1995). However, based on the present, more extensive geochemical and isotopic evaluation, it appears that the rocks of Group 3 are in fact composed of Archean gneisses with subordinate amounts of juvenile melt, which were rendered indistinguishable from Early Proterozoic magmatic rocks by thorough partial melting and recrystallization during Early Proterozoic deformation and metamorphism.

The chemical and isotopic characteristics of samples in Group 3 indicate they were derived largely from reworked Archean crust with Nd and Sr isotopic compositions similar to those of Nain Province tonalitic crust (see Chapter 2). εNd(M) and $^{87}Sr/^{86}Sr(M)$ values overlap with those of Nain Province gneisses (Figs. 13B and 14B). The REE patterns for Group 3 rocks are strongly fractionated, with $La_{(N)}/Yb_{(N)}$ between ca. 80 and 160 (Fig. 7A). In addition, several samples have large positive Eu anomalies (Figs. 7E and 12A). These patterns are identical to those of Archean gneisses in the Nain and Rae provinces, but are distinct from all other Early Proterozoic metaigneous rocks. Pb isotopic compositions overlap with those of Group 2 and probably reflect high mobility of U and Pb during mixing between Archean rocks and Proterozoic magmas, either at the time of Proterozoic magmatism or during subsequent metamorphism.

3.6.5 Origin of the Early Proterozoic magmas

On the basis of isotopic and trace element characteristics discussed above, it was concluded that the calc-alkaline metaigneous rocks comprising Groups 1 and 2 represent continental arc magmas which formed through fractional crystallization of a mantle-derived basalt. Direct partial melting of a mantle source cannot produce magmas any more felsic than basaltic andesite (Green and Ringwood, 1968), thus in order to produce the range of compositions observed in this study, a mantle-derived basaltic melt must undergo further fractionation. The rocks do not represent continuous crystal fractionation and separation from a single batch of magma, but reflect varying degrees and stages of crystal settling and magma separation. Many of the high-SiO_2 rocks show large Eu anomalies coupled with ca. 70 vol.% feldspar, indicating feldspar accumulation (e.g., LC92-12, -109). Other samples (e.g., LC92-40) appear to have formed through crystallization of relatively primitive magmas without significant fractionation.

3.6.5.1 Modeling of Group 1 - Fractional Crystallization

The geochemical, isotopic and field characteristics of Group 1 rocks imply that they formed from a common source and are related by igneous processes through which the observed compositional range (gabbro through trondhjemite) evolved. Below, we evaluate whether Group 1 samples are related by simple closed-system fractionation. Fig. 18 shows the evolution of Nd and Sr concentrations in response to closed-system fractional crystallization of a basaltic source using gabbro diorite sample LC92-40 as a proxy for the basalt. In reality, the true basaltic source probably had lower Nd and Sr concentrations, as LC92-40 does not represent a primary mantle melt (discussed above). A number of bulk distribution coefficients were used (Henderson, 1984; Reiners et al., 1996), and those illustrated show the geologically most reasonable approximations which fit Group 1 samples, assuming crystal fractionation from gabbro-dioritic to tonalitic melts. The data in Fig. 18 clearly show that the samples in Group 1 are not related to each other by simple, closed-system fractional crystallization. This does not, however, preclude an origin from a common source through fractional crystallization processes. Whereas the rocks may not have formed via a single closed system magma chamber, they could have evolved in isolated batches which varied in a number of parameters including crustal depth, temperature, bulk composition, intracrustal differentiation, and small amounts of crustal contamination. All of these factors likely affected the Kd-values and resulting element concentrations to such a degree that simple modeling (assuming closed system behavior during fractional crystallization), is inadequate to simulate the complexities and variations between magma batches.

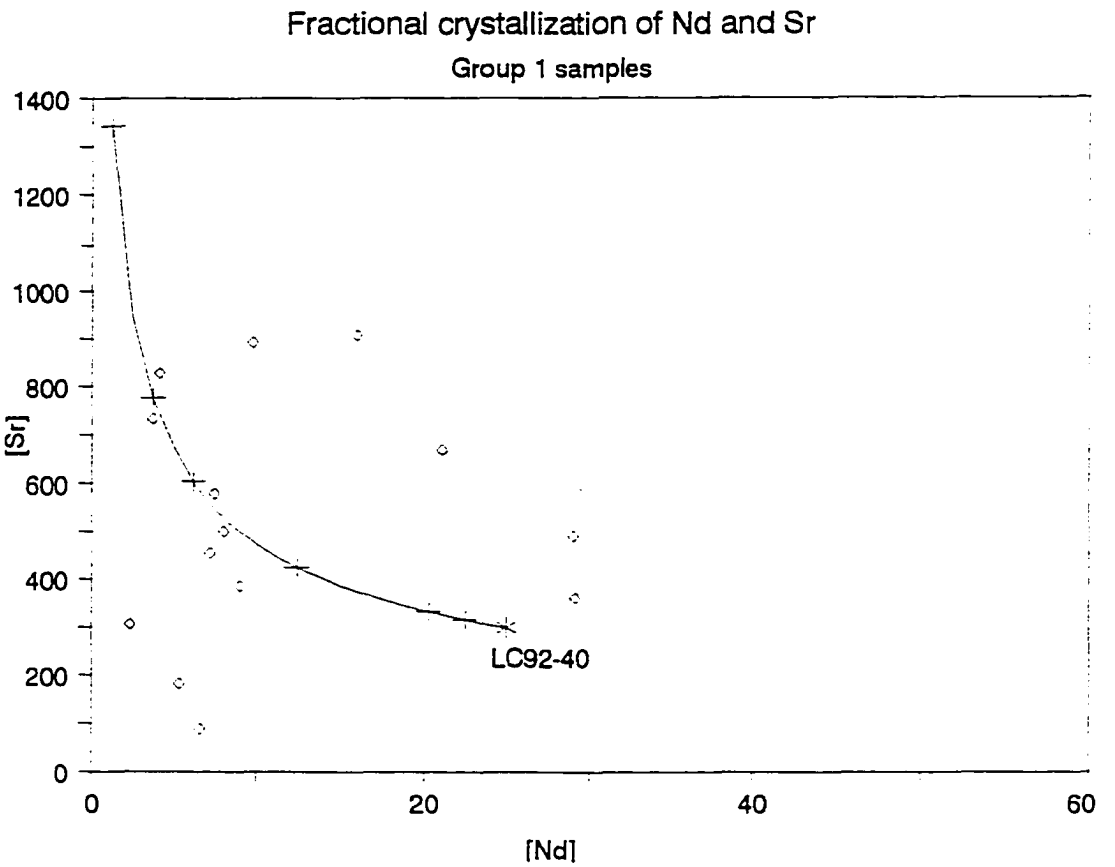


Figure 18. Results of fractional crystallization of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Calculations for fractional crystallization based on equations in Rollinson (1993). Nd (ppm) vs. Sr (ppm) showing the result of fractional crystallization using LC92-40 as the primary melt composition.

3.6.5.2 AFC modeling of Groups 1-3

The samples in Groups 1 and 2 show no correlation between SiO_2 and decreasing $\varepsilon Nd_{(1895)}$ or increasing $^{87}Sr/^{86}Sr_{(1895)}$ (Figs. 19A and B), hence they are not related through assimilation coupled with fractional crystallization (AFC). Nonetheless, results of AFC modeling are presented here to illustrate the effects that assimilation of local basement gneisses may have had on the Nd and Sr concentrations and isotopic compositions of rocks in Groups 1 and 2.

At the current level of exposure, the basement gneisses are laterally heterogeneous, but it is not clear whether the underlying crust is equally heterogeneous, or dominated by a particular lithology such as tonalitic gneisses or large volumes of underthrust/downfolded metasediments. We therefore use the two most locally abundant lithologies (i.e., Archean tonalitic basement gneiss and Tasiuyak paragneiss) as possible assimilants in an effort to simulate Nd and Sr concentrations and isotopic compositions for Groups 1, 2 and 3.

Group 1

Using the AFC equations of DePaolo (1981), Group 1 samples can be modeled using very low degrees of assimilation of Archean tonalitic crust (Figs. 20A and B). In these figures "r" represents the ratio between the mass of assimilated basement (Ma) and the mass that is fractionally crystallized (Mc). In Figs. 20A and B a curve with r = 0.005 is shown (i.e., a very small degree of assimilation relative to fractional crystallization) for comparison with Group 1 samples. Samples plot near the curve, suggesting that the Group 1 samples can be modeled by assimilating small but varying amounts of Archean tonalitic gneisses.

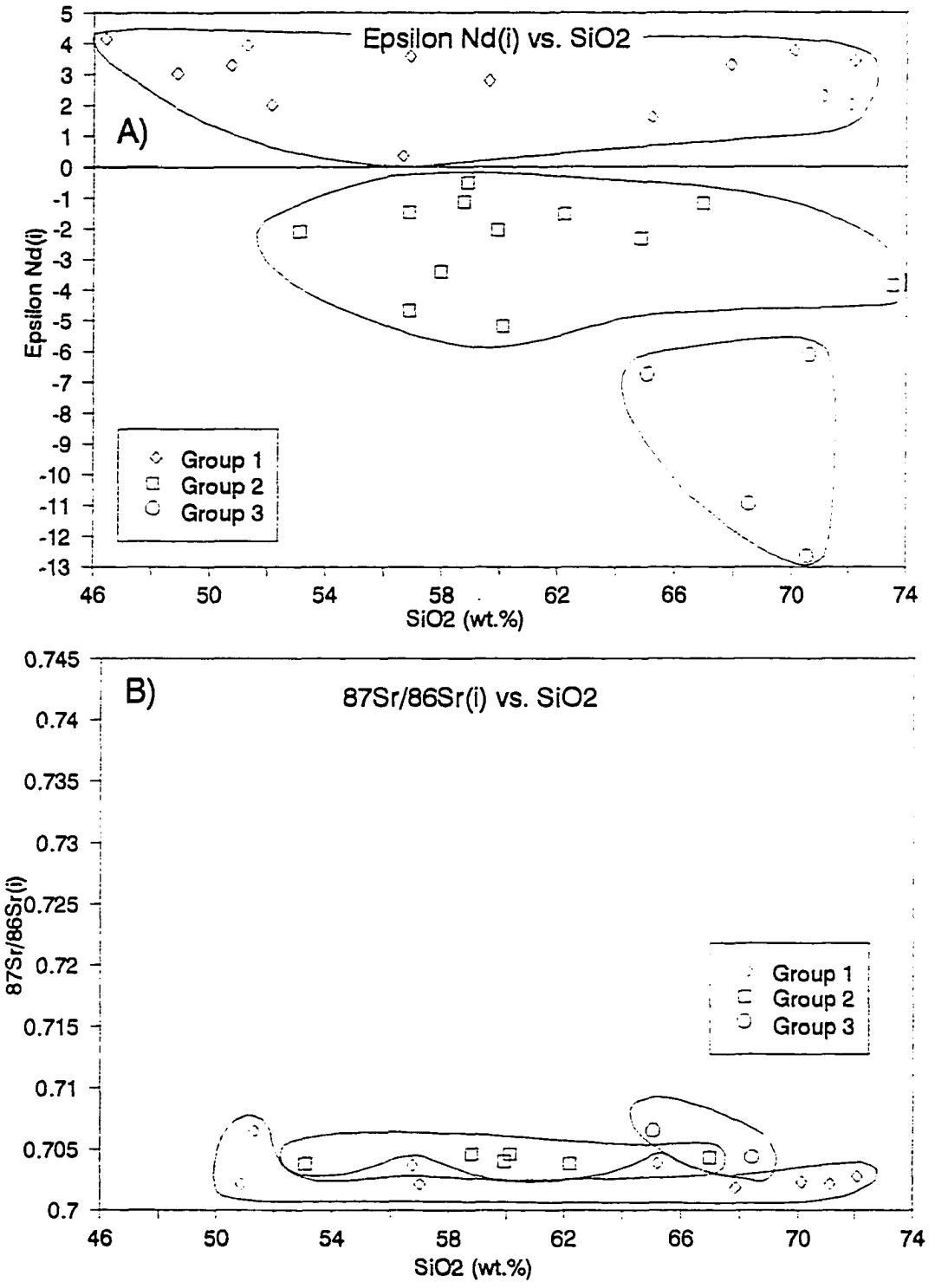


Figure 19. Initial Nd and Sr isotope composition vs. SiO_2 for Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. **(A)** - Initial εNd vs. wt% SiO_2. **(B)** - Initial $^{87}Sr/^{86}Sr$ vs. wt% SiO_2.

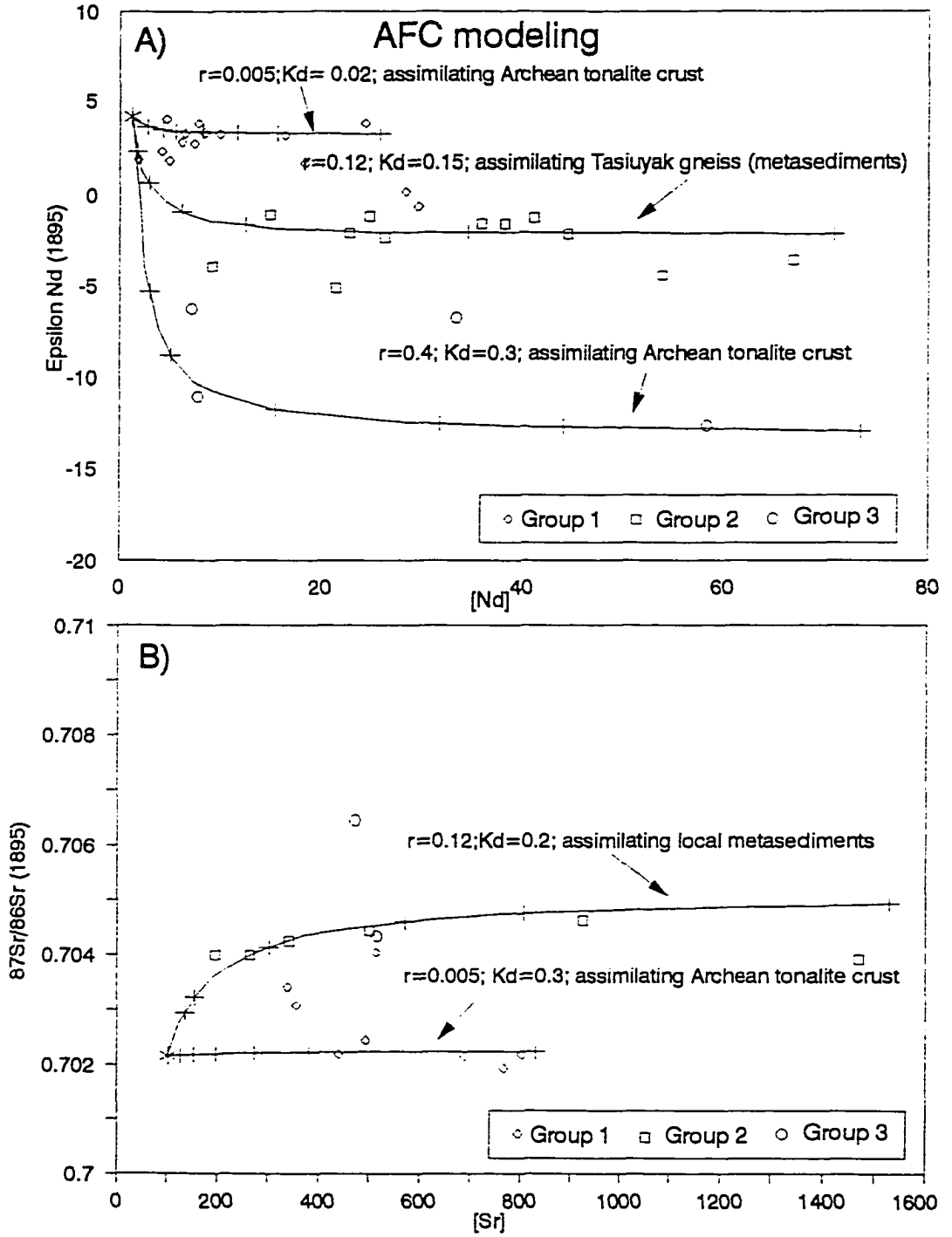


Figure 20. Results of AFC modeling of Early Proterozoic metaigneous rocks from Torngat Orogen, Northern Labrador. Calculations for AFC modeling based on expressions in DePaolo (1981). **(A)** - Nd (ppm) vs. εNd_{1895}, showing the results of varying proportions of assimilants (Archean tonalitic gneisses and metasedimentry Tasiuyak gneiss). **(B)** - Sr (ppm) vs. $^{87}Sr/^{86}Sr_{1895}$, showing selected values of "r" for different assimilants (Archean tonalitic gneiss and metasedimentary Tasiuyak gneiss). In (B) and (C), "r" is equal to Ma/Mc, where Ma = mass of the assimilant, and Mc = mass of fractional crystallization. "Kd" is the Nd and Sr bulk distribution coefficients (Arth, 1976; Reiners et al., 1996).

Group 2

Group 2 samples with low Nd concentrations can be modeled by slightly higher amounts of assimilated Archean tonalitic crust (r = 0.045; Kd_{Nd} = ca. 0.06; not shown). However, the r-curves corresponding to these parameters do not approach the high Nd concentrations observed among many Group 2 samples. The compositions of Group 2 samples are better approximated by assimilating larger amounts (r = 0.12) of the locally abundant metasedimentary Tasiuyak gneiss ($\varepsilon Nd_{(1895)}$ = -3.87, and Nd concentration ([Nd]) = 30.62; Figs. 19B and 19C).

Group 3

Nd concentrations and isotopic compositions of Group 3 samples can be approximated by using moderate assimilation (r = 0.12-0.4) of Archean tonalitic gneiss (Fig. 20A; r = 0.4 is shown). It is not possible to generate the high $^{87}Sr/^{86}Sr$ values observed in some Group 3 samples by using the same Archean tonalite as an assimilant. This problem is most likely the result of open system behavior of Rb and Sr during strong Proterozoic reworking (i.e., strong deformation and partial melting) of Group 3 samples.

In summary, Group 1 samples can be modeled using very small amounts of assimilated Archean crust. Group 2 samples with low Nd concentrations can be modeled using Archean tonalites as a crustal assimilant (not shown), but all Group 2 samples can be modeled by assimilating small amounts of metasediment (Tasiuyak gneiss). The Nd characteristics of Group 3 samples can be modeled by assimilating large amounts of Archean tonalitic crust, but modeling of Sr is unsuccessful, probably because of open system behavior of Rb and Sr during Proterozoic reworking of Group 3 rocks.

It has been shown that the rocks comprising the three groups are not directly related by AFC-processes and the present modeling experiments have furthermore shown

that the three groups are not related by assimilating varying amounts of a similar crustal source. In addition, it is interesting to note that while the compositions of Group 2 samples cannot be modeled by assimilating the local Archean orthogneisses, they can be modeled by assimilating the locally abundant metasediments.

3.6.6 Significance of geographic distribution of $\varepsilon Nd_{(1895)}$ values

The $\varepsilon Nd_{(1895)}$ values for the Early Proterozoic magmatic rocks show a very distinct geographic distribution (Fig. 3). Group 1 samples define an elongate core and arranged concentrically around this core are samples from Group 2. Group 3 samples occur only in the western part of the Burwell domain, where $\varepsilon Nd_{(1895)}$ values approach those of neighboring Nain and Rae Province Archean gneisses.

As shown previously, these magmatic rocks have trace element compositions which are indistinguishable from Phanerozoic arc rocks (Fig. 9) and are therefore assumed to have formed in arc-settings in response to subduction.

In northern Labrador, the Early Proterozoic metaigneous rocks intrude Nain Province gneisses on both sides of the Komaktorvik shear zone and are interpreted to have been generated during subduction beneath the northern Nain Province margin. If eastward dipping subduction is assumed then the spatial distribution of εNd values for the Proterozoic rocks does not conform to that typically observed in Phanerozoic continental margins (e.g., Condie and Potts, 1969; Farmer and DePaolo, 1983; DePaolo and Farmer, 1983).

In the following we present three mechanisms which may (alone or in concert) have contributed to the unusual distribution of εNd values.

The first model proposes that the area in which Group 1 rocks occur contains a thick pile of underplated basaltic material. This basalt could have been emplaced during a phase of (incipient?) rifting along the western margin of the Nain Province, presently

marked by the KSZ and the eastern margin of the Burwell domain (Fig. 21). The effect of this underplating is to substantially dilute the low-εNd lower crust with high-εNd material at this location. Subsequent arc magmatism which penetrated the lower crust may have triggered partial melting of the underplated basaltic material to generate the Group 1 magmas. A number of pieces of indirect evidence support this model, which involves the introduction of basaltic material in the lower crust:

(a) the occurrence of 2.4-2.2 Ga mafic dykes in the Nain Province on either side of the KSZ indicates a phase of mafic magmatism prior to convergence and collision.

(b) the restricted occurrence of anorthosites and interlayered amphibolites within the KSZ. The anorthosites most likely formed by magmatic differentiation of the underplated basalts, and were tectonically emplaced as continuous sheets into their present location in the KSZ by thrusting associated with Nain-Rae collision (Fig. 21C).

(c) the anorthosites and amphibolites are cut by Early Proterozoic magmas, so they are known to be older than 1910 Ma, the age of the oldest Early Proterozoic intrusive. Therefore the anorthosites and amphibolites were in place prior to emplacement of the Group 1 magmas.

Another way to explain the observed distribution of εNd values for the Early Proterozoic rocks across the northern TO is that the volume of Proterozoic magma was variable from east to west and that the region where Group 1 rocks occur contained the highest volume (Fig. 22). The effect of this would be to dilute the amount of assimilated crust such that Group 1 rocks would record less contamination than the other groups. There is, however, no supporting field evidence, which indicates that the volume of Proterozoic magmatism is higher in the region of Group 1 rocks than in areas where

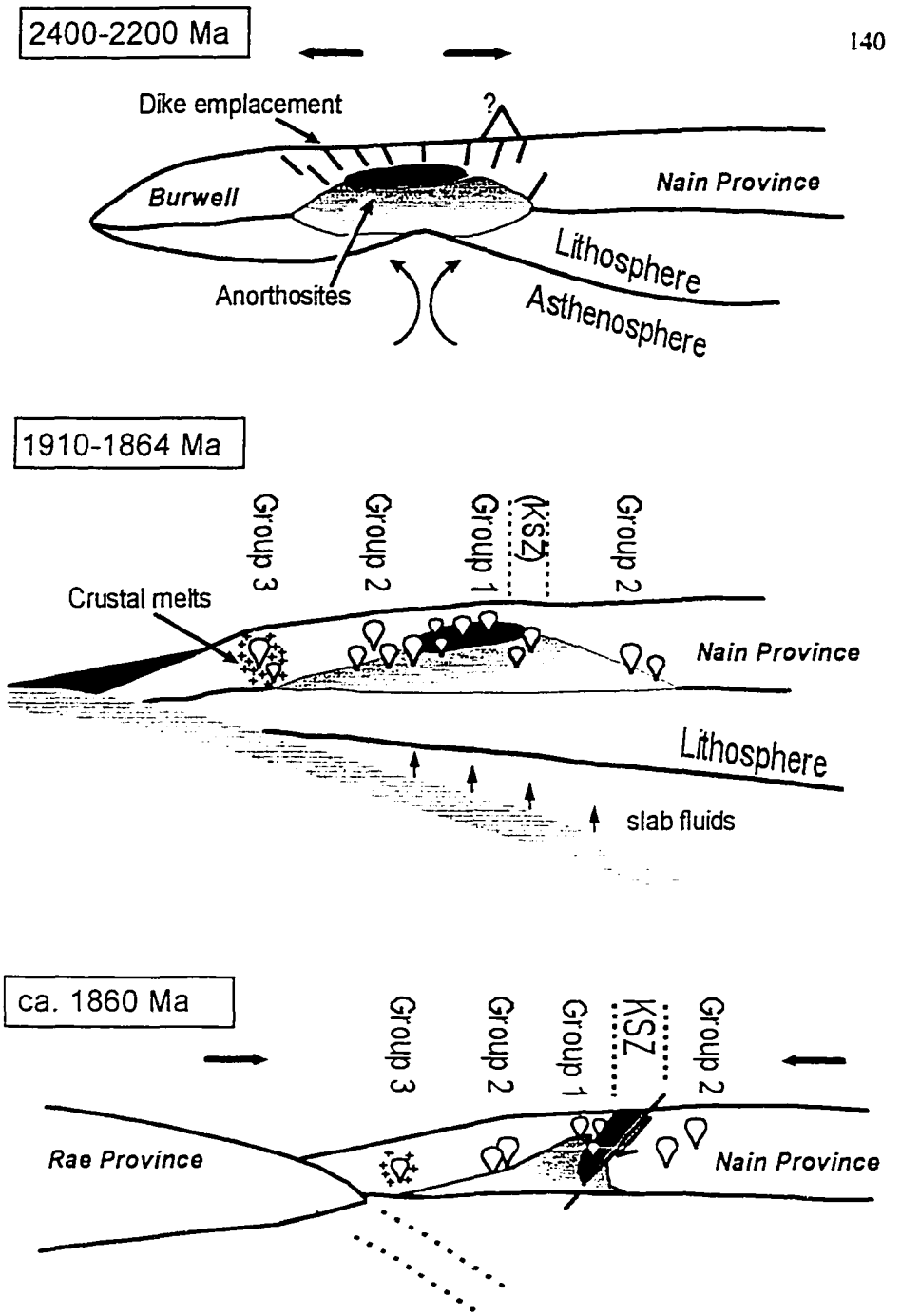


Figure 21. Cartoon cross-sections illustrating the proposed Early Proterozoic evolution in northern Labrador. (A): Basaltic underplating in the Nain Province margin in the period 2400-2200Ma (possibly accompanied by rifting) and associated dyke-emplacement and volcanism. Differentiation of the basaltic magma produces anorthosites. (B): During convergence between Nain and Rae provinces, subduction-generated magmas intrude the Nain margin in the period 1910-1864 Ma. Plutons emplaced through mafic underplated material record little or no contamination from continental crust (Group 1), whereas magmas emplaced outside or in the margin of the mafic underplated material record moderate to strong crustal contamination. The approximate location of the future Komaktorvik shear zone (KSZ) is indicated. (C): Anorthosites and amphibolites are emplaced as sheets within the KSZ by thrusting during Nain-Rae collision at ca. 1860 Ma.

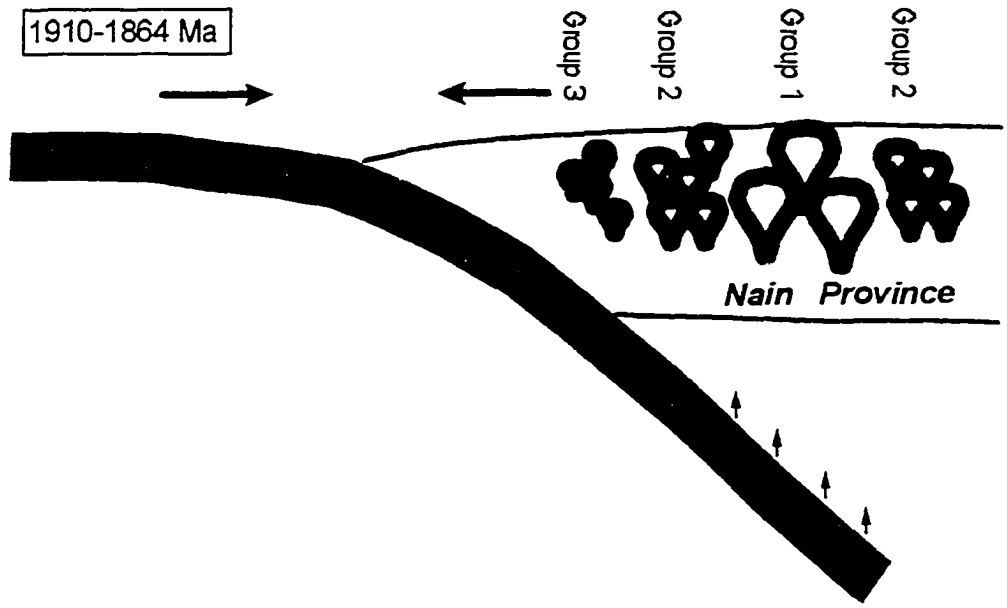


Figure 22. Schematic cross-section showing variations in the volume of arc magmatism across the northern T.O.. By volume, the larger plutons interact comparatively less with host rocks (represented by gray zones) than smaller plutons. Thus Group 1 magmas are relatively uncontaminated by the Archean basement, whereas groups 2 and 3 are more strongly contaminated by Archean basement gneisses.

Group 2 rocks occur, and therefore this mechanism is less likely to be solely responsible for the spatial variations in εNd values between Groups 1 and 2.

The third scenario which could explain to the spatial distribution of εNd values across the northern TO is that the KSZ marks a fundamental structural break in the Nain Province crust, and that the Burwell domain was thrust eastward along the KSZ and tilted east-side-up, so that the eastern Burwell domain exposes a deeper crustal level than the western Burwell (Fig. 23). With this geometry, exposed Group 1 magmas occur lower in the crust and therefore had less chance of interaction with the Archean basement gneisses. In contrast, Group 3 rocks would represent shallow level crustal melts where melting was triggered by magma emplacement deeper in the crust. There is, however, no current metamorphic data to support this sort of metamorphic gradient and thus crustal tilting of this magnitude (e.g., Mengel & Rivers, 1997; Campbell, unpubl. data).

3.7 CONCLUSIONS

Geochemical and Nd, Sr and Pb isotope data for calc-alkaline metaigneous rocks occurring within the Burwell domain and the Nain Province in Northern Labrador have provided valuable constraints for the origin of the metaigneous rocks and the tectonic development of the northern Torngat Orogen.

Geochemical characteristics of the Early Proterozoic metaigneous rocks are similar to those of Phanerozoic arc magmas. This, coupled with the preservation of intrusive contacts between the Early Proterozic metaigneous rocks and Archean basement gneisses on either side of the KSZ is indicative of subduction-related formation in a continental arc. Eastward dipping subduction offers the only reasonable

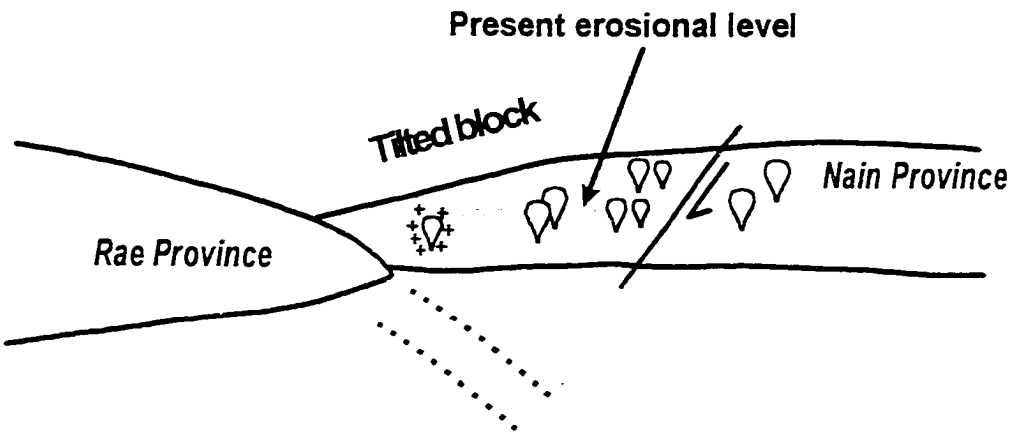


Figure 23. Schematic cross-section of the relation between crustal scale tilting and magma-host rock interaction in the northern Torngat Orogen. At the present erosion levels, the deepest levels are exposed in the eastern part of the tilted block. Assuming that magmas interact with their host rocks on their way up through the crust, magmas at lower crustal levels will record less contamination than those at higher crustal levels. In this way, the eastern plutons will be less contaminated, whereas the central and western ones will have experienced more interaction with crustal material.

model by which these arc magmas would have formed since they occur on either side of the KSZ, but only to the north and east of the Abloviak shear zone.

On the basis of $\varepsilon Nd_{(1895)}$ values, the Early Proterozoic metaigneous rocks define three groups. Group 1 samples ($\varepsilon Nd_{(1895)} > 0$) approximate depleted mantle values and are thus relatively uncontaminated by the basement gneisses into which they intrude. Group 2 samples ($-5 < \varepsilon Nd_{(1895)} < 0$) indicate substantial involvement of older crust. Group 3 samples ($\varepsilon Nd_{(1895)} < -5$) include a large to dominant component of Archean crust. The spatial distrubution of the Early Proterozoic metaigneous rocks is as follows: Group 1 rocks occur in an elongate, north-south trending core along the eastern margin of the Burwell domain. Group 2 rocks occur peripherally around Group 1, and Group 3 occurs only along the western margin of the Burwell domain.

This spatial distribution of $\varepsilon Nd_{(1895)}$ for the Early Proterozoic metaigneous rocks does not comply with a typical Phanerozoic-style eastward dipping subduction model, in which $\varepsilon Nd_{(1895)}$ values would decrease away from the subduction zone (eastward), reflecting an increasing interaction of the mantle derived magmas with Archean basement gneisses. We propose a new model, which accounts for the unusual distribution of $\varepsilon Nd_{(1895)}$ values as well as a number of other features in the northern TO. In this model, the area now marked by the KSZ and the eastern margin of the Burwell domain was underplated by a thick pile of basalt, possibly associated with rifting or incipient rifting. This event occurred prior to the the onset of convergence and may be related to the emplacemnt of ca. 2.4-2.2 Ga dike swarms which intrude Nain Province basement gneisses on either side of the KSZ. Beginning at ca. 1910 Ma, arc magmatism occurred across the area in response to convergence and eastward subduction beneath the Nain margin. The arc magmas which were intruded into the area underlain by a thick pile of basalts did not interact with or assimilate a significant amount of Archean basement gneisses, thus the resulting Group 1 samples had mantle-like $\varepsilon Nd_{(1895)}$ values.

Alternatively, arc magmatism may have caused partial melting of the underplated basalts to produce Group 1 samples. Subsequently, sheets of anorthosites and interlayered amphibolites were emplaced into their current position in the KSZ by thrusting associated with Nain-Rae collision.

REFERENCES

Barbey, P, Cuney, M., 1982. K, Rb, Sr, Ba, U and Th geochemistry of Lapland granulite (Fennoscandia) LILE fractionation controlling factor. Contrib. Mineral. Petrol., 81: 304-316.

Bertrand, J.-M., Roddick, J.C., Van Kranendonk, M.J., Ermanovics, I. 1993. U-Pb geochronology of deformation and metamorphism across a central transect of the Early Proterozoic Torngat Orogen, North River map area, Labrador. Can. J. Earth Sci., 30: 1470-1489.

Collerson, K.D., Campbell, L.M., Weaver, B.L., Palacz, Z.A., 1991. Evidence for Extreme Mantle Fractionation in Early Archean Ultramafic Rocks from Northern Labrador. Nature, 349: 209-214.

Connelly, J. N., Mengel, F. C., 1994. Preliminary U-Pb data from Archean gneisses in a transect across the Komaktorvik Shear Zone, Northern Torngat Orogen. *In* ESCOOT Lithoprobe Report 45; 72-82.

Condie, K.C., Potts, M.J., 1969. Calcalkaline volcanism and the thickness of the early Precambrian crust in North America,. Can J. Earth Sci., 6: 1179-1184.

Condie, K.C., 1997. Contrasting Sources for Upper and Lower Continental Crust: The Greenstone Connection. Jour. Geol., 105:729-736.

DePaolo, D.J., 1981. Neodymium isotopes in the Colorado Front Range and crust-mantle evolution in the Proterozoic. Nature: 291/193-196.

DePaolo, D.J., 1988. Neodymium Isotope Geochemistry. Springer-Verlag, Pubs., 187pp.

DePaolo, D.J., Farmer, G.L., 1983. Isotopic data bearing on the origin of Mesozoic and Tertiary granitic rocks in the western United States.*In* The relative contribution of mantle, oceanic crust and continental crust to magma genesis. *Edited by* S. Moorbath, R.N. Thompson and E.R. Oxburgh. Phil. Trans. R. Soc. Lond. A 310, 305-316.

Ermanovics, I., 1993. Geology of the Hopedale Block, southern Nain Province, and the adjacent Proterozoic terranes, Labrador, Newfoundland. Geological Survey of Canada, Memoir 431, 161p.

Ermanovics, I., Van Kranendonk, M.J., 1990. The Torngat Orogen in the North River-Nutak transect area of Nain and Churchill provinces. Geosci. Can., 17: 293-296.

Farmer, G.L., DePaolo, D.J., 1983. Origin of Mesozoic and Tertiary granite in the western United States and Implications for Pre-Mesozoic crustal structure 1. Nd and Sr isotopic studies in the geocline of the northern Great Basin. Jour. Geophys. Res.; 88: 3379-3401.

Green, T.H, Ringwood, A.E., 1968. Genesis of the calc-alkaline igneous rock suite. Contrib. Mineral. Petrol. 18: 105-162.

Heier, K.S. 1965. Radioactive elements in the continental crust. Nature, 208: 479-480.

Heier, K.S. 1973. Geochemistry of granulite facies rocks and problems of their origin. Phil. Trans. R. Soc. Lond. A273: 429-442.

Henderson, P., 1984. Rare Earth Element Geochemistry. Elsevier Publishers. 510pp.

Hildreth, W., Moorbath, S., 1988. Crustal contributions to arc magmatism in the Andes of Central Chile. Contrib. Mineral. Pet. 98: 455-489.

Hoffman. P.F., 1988. United Plates of America, the birth of a craton: Early Proterozoic assembly and growth of Laurentia. Ann Rev. Earth Planet. Sci.; 16: 543-603.

Korstgaard, J., Ryan, B., Wardle, R. 1987. The boundary between Proterozoic and Archean crustal blocks in central west Greenland and northern Labrador, In Evolution of the Lewisian and Comparable Precambrian High Grade Terrains, *Eds.* R.G. Park, J. Tarney. Geol. Soc. Sec. Pub. 27: 247-259.

McCulloch, M.T., Gamble, J.A., 1990. Geochemical and geodynamical constraints on subduction zone magmatism. Earth Planet. Sci. Let., 102: 358-374.

Mengel, F., 1988. Thermotectonic evolution of the Proterozoic-Archaean boundary in the Saglek area, northern Labrador. Ph. D. thesis, Memorial University of Newfoundland, 335 p.

Pearce, J.A., 1983. The role of sub-continental lithosphere in magma genesis at destructive plate margins. *In* Continental basalts and mantle xenoliths. C.J. Hawkesworth and M.J. Norry, *Eds.* Nantwich: Shiva *Pubs.*: 230-249.

Reiners, P.W., Nelson, B.K., Nelson, S.W., 1996. Evidence for Multiple Mechanisms of Crustal Contamination of Magma from Compostionally Zoned Plutons and Associated Ultramafic Intrusions of the Alaska Range. Jour. Petrol., 37/2:261-292.

Rivers, T., Mengel, F., Scott, D.J., Campbell, L.M., Goulet, N., 1996. Torngat Orogen-a Paleoproterozoic example of a narrow doubly vergent collisional orogen. *In* Precambrian Crustal Evolution in the North Atlantic Region, *Ed.* T.S. Brewer, Geol. Soc. Spec. Pub. 112: 117-136.

Rudnick, R.L., McLennan, S.M., Taylor, S.R., 1985. Large ion lithophile elements in rocks from high-pressure granulite facies terrains. Geochim. Cosmochim. Acta; 49: 1645-1655.

Scott, D.J., 1995a. U-Pb geochronology of a Paleoproterozoic continental magmatic arc on the western margin of the Archean Nain craton, northern Labrador, Canada. Can. J. Earth Sci., 32/11: 1870-1882.

Scott, D.J., 1995b. U-Pb geochronology of the Nain craton on the eastern margin of the Torngat Orogen, Labrador. Can. J. Earth Sci., 32/11: 1859-1869.

Scott, D., Campbell, L.M., 1993. Evolution of the Paleoproterozoic Torngat Orogen, Labrador Canada: Recent advances using U-Pb Geochronology and Nd isotopic systematics. Geological Society of America, Abstracts with Program, vol. 25, p. A-237.

Scott, D.J., Machado, N., 1994. U-Pb geochronology of the northern Torngat Orogen, Labrador, Canada: a record of Paleoproterozoic magmatism and deformation. Precambrian Res., 70: 169-190.

Scott, D. J., Machado, N., 1995. U-Pb geochronology of the northern Torngat Orogen, Labrador, Canada: a record of Paleoproterozoic magmatism and deformation. Precambrian Res. 70: 169-190.

Shaw, D.M., 1968. A review of K/Rb fractionation trends by covariance analysis. Geochim. Cosmochim. Acta. 32: 573-601.

Stacey, J. S., Kramers, J. D. (1975) Approximation of terrestrial lead isotope evolution by a two-stage model. Earth Planet. Sci. Lett., 26: 207-221.

Stern, R.A., Percival, J.A., Mortensen, J.K., 1994. Geochemical evolution of the Minto block: a 2.7 Ga continental magmatic arc built on the Superior proto-craton. Precambrian Res., 65: 115-153.

Sun, S. S., 1980. Lead isotopic study of young volcanic rocks from mid-ocean ridges, ocean islands and island arcs. Phil. Trans. Royal Soc. London, A297: 409-445.

Tatsumi, Y., 1989. Migration of fluid phases and genesis of basalt magmas in subduction zones. J. Geophys. Res., 94: 4697-4707.

Taylor, F. C., 1974. *In* Age determinations and geological studies, K-Ar isotopic ages, Report 12. Geological Survey of Canada, Paper 74-2.

Theriault, R.J., Ermanovics, I., 1996. Sm-Nd isotopic and geochemical characterization of the Paleoproterozoic Torngat orogen, Labrador, Canada. Precambrian Res., 81: 15-35.

Touret, J.L.R., 1996. LILE-depletion in granulites: Myth or reality? *In* Petrology and Geochemistry of magmatic suites of rocks in the continental and oceanic crusts. A volume dedicated to Professor Jean Michot *Ed.* D Demaiffe: 53-72.

Van Kranendonk, M.J., Ermanovics, I., 1990. Structural Evolution of the Hudsonian Torngat Orogen in the North River map Area, Labrador: evidence for east-west transpressive collision of Nain and Rae continental blocks. Geosci. Can., 17: 283-288.

Van Kranendonk, M.J., Scott, D.J., 1992. Preliminary report on the geology and structural evolution of the Komaktorvik Zone of the Early Proterozoic Torngat Orogen, Eclipse Harbour area, northern Labrador. Current Research, Part C; Geological Survey of Canada, Paper 92-1C: 59-68.

Van Kranendonk, M.J., Wardle, R. J., Mengel, F., Ryan, B., Rivers, T., 1993. Lithotectonic divisions of the northern part of the Torngat Orogen, Labrador, Quebec and N. W. T. *In* ECSOOT Report of Transect meeting. Eds. R. J. Wardle , J. Hall. Lithoprobe Report 32: 21-31.

Van Kranendonk, M.J., Wardle, R.J., 1996. Burwell domain of the Paleoproterozoic Torngat Orogen, northeastern Canada: Tilted cross-section of a magmatic arc caught between a rock and a hard place. *In* Precambrian Crustal Evolution in the North Atlantic Region, *Ed.*T.S. Brewer, Geol. Soc. Spec. Pub. 112: 91-116.

Van Kranendonk, M.J., Wardle, R.J., *in press*. Crustal scale flexural slip folding during late tectonic amplification of an orogenic boundary perturbation in the Paleoproterozoic Torngat Orogen, NE Canada. Can. J. Earth Sci. (1997)

Wardle, R. J., 1983. Nain-Churchill province cross-section, Nachvak fiord, Northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 83-1: 68-89.

Wardle, R.J., Van Kranendonk M.J., Mengel, F., Scott, D., 1992. Geological mapping in the Torngat Orogen, Northernmost Labrador: Preliminary Results. *In* Current Research, Newfoundland Dept. of Mines and Energy, Geological Survey Branch, Report 92-1: 413-429.

Wilson, M., 1989. Igneous Petrogenesis. Unwin Hyman Pub., 466pp.

Wyllie, P.J., Sekine, T., 1982. The formation of mantle phlogopite in subduction zone hybridization. Contrib. Mineral. Petrol. 79: 375-380.

Chapter 4

Models for the growth of Laurentia: revisions based on investigations in the Torngat Orogen, northern Labrador, Canada

4.1 Introduction

Models for the formation and development of continents during the Precambrian have changed radically duing the past 30 years due to the application of Phanerozoic style plate tectonic processes as a guide to Precambrian tectonic development. Hoffman (1988) elegantly provided the first detailed synthesis of the Early Proterozoic amalgamation of Archean crustal blocks which lead to the construction of Laurentia, the North American craton. In this work, Hoffman describes the formation of Laurentia as the result of interaction among previously independent Archean crustal fragments along a network of collision zones, or orogenic belts. Hoffman employed the mechanisms of modern plate tectonics to describe the process which drove the construction of Laurentia. This work provided a geological framework in which individual crustal provinces and adjoining collision belts (orogens) could be successively viewed and interpreted.

The unifying work by Hofmann (1988) paved the way for a number of other studies, which used similar methodologies to provide syntheses of Precambrian terranes elsewhere in the world (e.g., Baltica: Gaal and Gorbatschev, 1987; Gorbatschev and Bogdanova, 1993; North Atlantic Craton: Bridgwater et al., 1990, 1991; Park, 1994, 1997). New data allowed Hoffman (1990), followed by Van Kranendonk et al. (1993), to develop updated syntheses and models for the assembly of northeastern Laurentia.

On the basis of these syntheses and models of Early Proterozoic orogenesis, it has become gernerally accepted that the welding of Archean cratons along Early Proterozoic orogenic belts was driven by plate tectonic processes which broadly resemble more modern, e.g., Phanerozoic, orogens. This convergence of thinking can be attributed to the preservation and successful identification of

geological features such as crustal scale sutures marked by obducted ophiolite complexes and accretionary wedge sediments, accreted island arcs, continental arcs and rifted passive margins.

This chapter reviews the models by Hoffman (1988, 1990) and Van Kranendonk et al. (1993), which discuss the Early Proterozoic assembly of NE Laurentia. The current models for Proterozoic development of the Torngat Orogen are also presented, followed by a review of the results presented in Chapters 2 and 3 of this thesis. The results of this work require modifications to the existing models, and the resulting implications for the geometry and growth of Laurentia during the Early Proterozoic are presented and discussed.

4.2 Early Proterozoic assembly and growth of Laurentia: the model of Hoffman (1988)

Laurentia was originally defined as an amalgam of Archean cratons and intervening Proterozoic orogens which comprise what is today North America and southern Greenland (Hoffman, 1988; Fig. 1). The Archean cratons, or provinces, making up large parts of Laurentia are the Slave, Churchill, Wyoming, Superior and Nain provinces. The Churchill Province is further subdivided into an Archean component: the Hearne and Rae Provinces, and an Early Proterozoic component: the Trans-Hudson orogen (THO). The Rae Province comprises a northern and a southeastern arm (N Rae and SE Rae, respectively). The Archean provinces are composed of granite-greenstone terrains (e.g., Slave and Hearne Provinces), and/or granulite facies gneiss terrains (e.g., Nain Province), each of which include varying amounts of intercalated supracrustal material. The cratons are overlain by the erosional remnants of Early Proterozoic supracrustal rocks, and both the cratons and their cover are variably metamorphosed and deformed in response to the Early Proterozoic orogeneses that occurred along the cratonic margins during assembly.

In Hoffman's (1988) synthesis, the northeastern part of Laurentia was composed of the Archean Superior, southeast Rae, and Nain provinces (Fig. 1). The Early Proterozoic collisional belts which border the Archean provinces are typically asymmetric, with one side dominated by a sedimentary prism thrust toward the Archean foreland, and the other side characterized by a magmatic arc. The asymmetry is interpreted to reflect the polarity of subduction during ocean

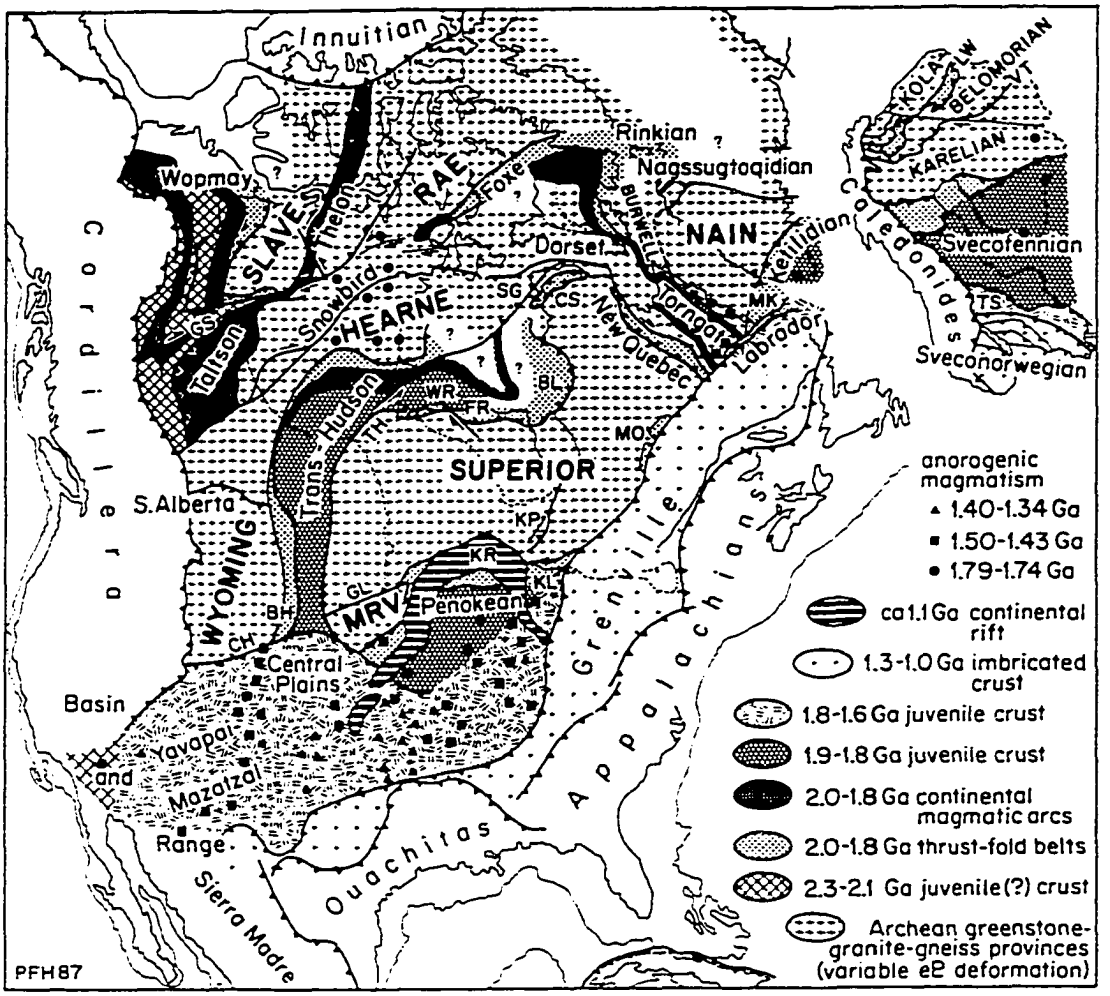


Figure 1. Precambrian tectonic elements of Laurentia from Hoffman, 1988. The Baltic shield is shown in a pre-Iapetus reconstruction., and Greenland is restored to rifting from North America. Uppercase names are Archean provinces; lowercase names are Proterozoic and Phanerozoic orogens. Abbreviations: BH, Black Hills inlier,; BL, Belcher belt; CH, Cheyenne belt; CS, Cape Smith belt, FR, Fox River belt; GL, Great Lakes tectonic zone; GS, Great Slave Lake shear zone; KL, Killarney magmatic zone; KP, Kapuskasing uplift; KR, Keweenawan rift zone; LW, Lapland-White Sea tectonic zone; MK, Makkovikorogen; MO, Mistassini-Otish basins; MRV, Minnesota River Valley terrane; SG, Sugluk terrane; TH, Thompson belt; TS, Transscandinavian (Smaland-Varmland) magmatic zone; VT, Vetrenny tectonic zone; WR, Winisk River fault.

closure. The main geological features of these Early Proterozoic orogenic belts are briefly summarized below (following Hoffman, 1988). Geochronological data is referenced in Hoffman (1988):

4.2.1 Cape Smith Belt

The Cape Smith Belt (CS on Fig. 1; N,S and W on Fig 2) represents the erosional remnant of a thin-skinned south-vergent thrust belt. The volcanic and sedimentary rocks comprising the CSB record the evolution of a rifted northern margin of the Superior province from continental to transitional-oceanic to true oceanic crust, and includes a metamorphosed ophiolite suite which is intruded by ca. 1.84 Ga calc-alkaline magmatic rocks (St-Onge and Lucas, 1990; Scott et al., 1991, 1992.

4.2.2 New Quebec Orogen

The New Quebec orogen (New Quebec; Figs. 1 and 2) marks the suture between the Archean foreland of the Superior Province to the west and Archean hinterland of the Rae Province to the east. The Orogen is underlain by a complex collage of supracrustal sequences which were thrust eastward and metamorphosed in response to Rae/Superior collision between 1.87-1.79 Ga. East of the NQO, the ca. 1.84 Ga DePas batholith intrudes rocks of the Rae Province and is interpreted as a continental magmatic arc which formed in response to eastward subduction beneath the Rae Province. The southern part of the batholith and adjacent Archean gneisses are thrust westward on structures that parallel the Torngat Orogen. This deformation postdates the ca. 1.87-1.79 Ga thrusting in the New Quebec orogen, and is interpreted to suggest that the Torngat Orogen is younger than early phases of the New Quebec Orogen.

4.2.3 Foxe Fold Belt

The east-west trending, Early Proterozoic Foxe fold belt in Baffin Island (Foxe; Figs. 1 and 2) extends eastward into the Rinkian fold belt in central west Greenland (e.g. Grocott and Pulvertaft,

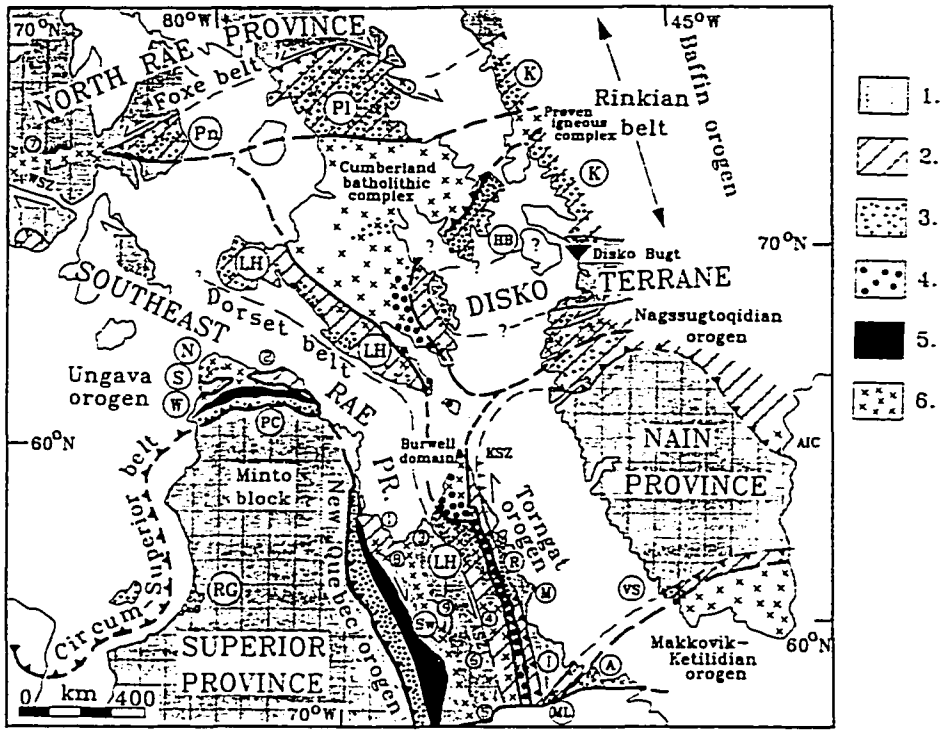


Figure 2. Geological compilation map of Northeastern Laurentia from Van Kranendonk et al., 1993, showing inferred distribution of Archean cratonic blocks and Early Proterozoic orogens. 1= areas of unreworked Archean crust; 2= areas of reworked crust; 3= Early Proterozoic shelf-turbidite sequences; 4=Tasiuyak gneiss; 5= mafic volcanics; 6=Early Proterozoic arc rocks. Early Proterozoic supracrustal assemblages are as follows: A=Aillik Group; HB= Hoare Bay Group; I= Ingrid Group; K=Karrat Group; LH= Lake Harbor Group; M= Mugford Group; ML=Moran Lake Group; N= Narsajuaq arc; PC= Povungnituk and Chukotat groups; Pl= Piling Group; Pn= Penrhyn Group; R= Ramah Group; RG= Richmond Gulf Group; S= Sugluk Group; Sw= Seward subgroup; VS= Vallen and Sortis groups; W=Watts Group. ASZ= Abloviak shear zone; KSZ= Komaktorvik shear zone; WSZ= Wager shear zone. AIC= Ammassalik igneous complex. Pre-drift reconstruction from Rowley and Lottes (1988).

1990). It is composed of Early Proterozoic platformal quartzite, schist and marble overlain by pelite and local mafic flows and graywacke turbidites resembling foredeep flysch (the Piling group). The supracrustal package has been shortened to produce tight, steeply inclined folds, and the rocks were subsequently refolded and variably metamorphosed at ca. 1.8 Ga. in response to transcurrent shearing.

4.2.4 Torngat Orogen: formed by Nain-Rae collision

The Torngat Orogen was described by Hoffman (1988) as the zone of intense Early Proterozoic deformation and metamorphism between the Rae and Nain provinces in northern-most Labrador, Canada (following Taylor, 1979; Ryan et al., 1983, 1984; Wardle, 1983, 1984; Mengel, 1984, 1985; Torngat, Figs. 1 and 2). As described in Chapter 1, collisional deformation between the two cratons was dominated by east-vergent thrusting in the Nain Province margin (representing the foreland), and sinistral transcurrent shearing in the margins of both provinces. Movement between the cratons was focused primarily along the Abloviak shear zone (ASZ), which coincides with a thick belt of garnetiferous mylonite developed from pelitic metasediments, the protoliths of which were thought to include off-shelf correlatives of Rae or Nain province cover sediments (Hoffman, 1988). The ASZ is a crustal scale suture, thought to bifurcate in the north around the Burwell terrain (Taylor, 1979, Korstgaard et al., 1987), a possible Archean microcontinent. At the point of bifurcation, the younger Komaktorvik shear zone (KSZ) trends northward and the ASZ bends westward. Collectively, the two shear zones frame the Burwell terrain. A magnetic low associated with the Tasiuyak gneiss continues offshore into Ungava Bay, where it bends to the north and can be traced as far as Resolution Island, off the southeast tip of Baffin Island.

4.3 Early Proterozoic assembly and growth of Northeastern Laurentia: the model of Hoffman (1990)

As a follow-up to his descriptive 1988 synthesis of the Early Proterozoic assembly of Laurentia, Hoffman (1990) proposed a model for the development of the northeastern part of

Laurentia, based on studies of the Cenozoic tectonic evolution of Southeast Asia (e.g., Peltzer and Tapponnier, 1988). In this model, the geometry of Archean cratonic fragments and the timing of collision between them in northern Labrador was suggested to result from northerly indentation of the Superior Province and consequent extrusion tectonics. The southeast Rae Province rotated clockwise in response to collisional indentation by the Superior Province (Fig. 3), which in turn caused the opening of the ancestral Baffin basin (into which supracrustal rocks of the Foxe, Rinkian and Dorset fold belts were deposited). Continued movement of the Superior indentor drove convergence between the Southeast Rae Province (hinterland) and Superior Province (foreland), followed by collision along the New Quebec orogen at ca. 1.88 Ga (Chevé and Machado, 1988). Subsequently, convergence between the Nain and Rae provinces was accommodated by subduction of oceanic crust beneath the eastern margin of the Rae Province. Subduction terminated in response to ca. 1.86 Ga docking of the Nain against the Rae Province, initiating Torngat orogenesis. In this model, the Burwell terrain represents an Archean crustal block which extends northward to include eastern Baffin island and West Greenland north of the Nagssugtoqidian orogen (Fig. 2). The Burwell terrain collided with the northwest margin of the Nain Province either before or after it collided with the Rae Province.

4.4 Early Proterozoic assembly and growth of Northeastern Laurentia: the model of Van Kranendonk et al. (1993) - differences with previous models

The model of Hoffman (1990) was proposed at a time when the northern segment of the TO was still largely Terra Incognita. Taylor's (1979) reconnaissance mapping project provided the only maps, and the ages of the component basement gneisses and the timing of orogenesis were very poorly constrained. A three-year field project in the northern Torngat region was initiated in 1991 jointly by the Newfoundland Department of Mines and Energy and the Geological Survey of Canada. As a result of this mapping project, much new data on structural, metamorphic, isotopic, geochemical and geochronologic aspects of the northern Torngat Orogen became available,

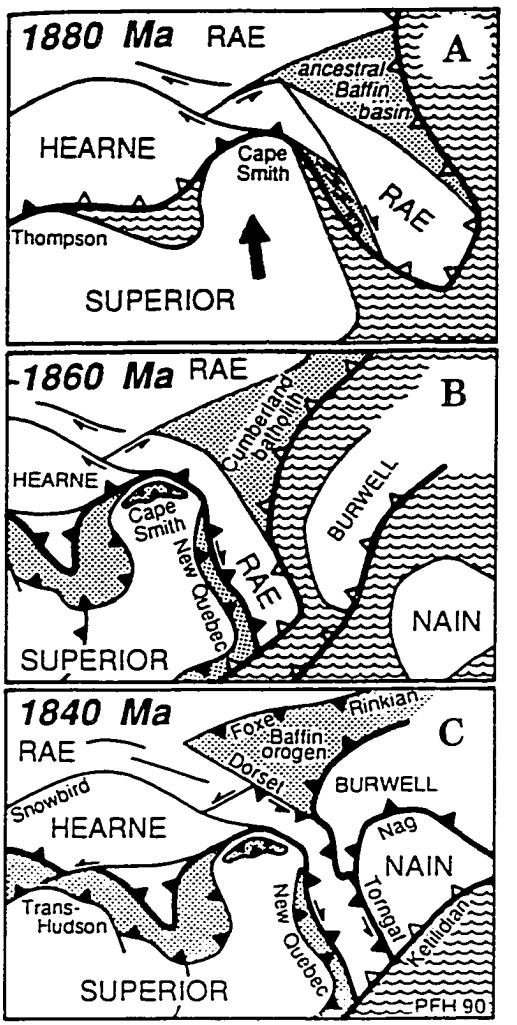


Figure 3. Hoffmans (1990) model for the tectonic assembly of Northeastern Laurentia. (A) Indentation of the Superior Province into Rae Province hinterland results in extrusion of an arm of the Rae Province to the southeast and formation of the ancestral Baffin basin. (B) Southerly thrusting of the Cape Smith beltin the Ungava orogenaccompanies collision between the southeast Rae and Superior provinces and emplacement of the Cumberland Batholith complex. (C) Accretion of Burwell and Nain Provinces. Modified according to Van Kranendonk et al., 1993, from Hoffman, 1990.

including:

1) the timing of collision along the Ungava, Torngat and New Quebec orogens (data from Perreault et al., 1988; Machado et al., 1988; Perreault and Hynes, 1990; Bertrand et al., 1993).

2) the timing of Early Proterozoic calc-alkaline magmatism in the central Torngat Orogen and in the Cumberland batholith (data from Bertrand et al., 1993; Pidgeon and Howie, 1975; Henderson, 1985; Jackson et al., 1990).

3) the ages and field characteristics of Archean gneisses occurring in the Torngat Orogen, Northern-most Labrador (data from Scott and Machado, 1994; Scott, 1995b).

Contrary to Hoffman, who proposed that the geometry of Archean crustal blocks resulted entirely from northerly indentation by the Superior Province and consequent extrusion tectonics, Van Kranendonk et al. (1993) suggested that indentation by the Superior Province was the last in a series of multiple collisional events between 1.92-1.74 Ga. According to this new model, the Burwell domain did not constitute an independent Archean craton unrelated to the Nain or southeast Rae Provinces, as proposed by Hoffman (1990), but consisted of Nain Province gneisses in the east, and Early Proterozoic charnockites in the west. Archean crust unrelated to the Nain Province was, however, argued to occur north of the Nagssugtoqidian Orogen in Central West Greenland and on Southeast Baffin Island. Collectively, this Archean crustal fragment was called the Disko terrane (Fig. 3) and was proposed as a northerly indentor which collided with the northern Rae Province and its overlying sedimentary cover to form the Foxe and Rinkian fold belts at >1876 Ma (age from Van Kranendonk et al., 1993; Fig. 4) along the Baffin Orogen.

Collision along the Baffin Orogen was preceded by westerly subduction of oceanic crust under the Southeast Rae Province. The resulting ca. 1950-1880 Ma magmatic arc suite is argued to comprise the Cumberland batholith complex and correlative charnockites that intrude the Tasiuyak gneiss and the eastern margin of the Southeast Rae Province in the Torngat orogen.

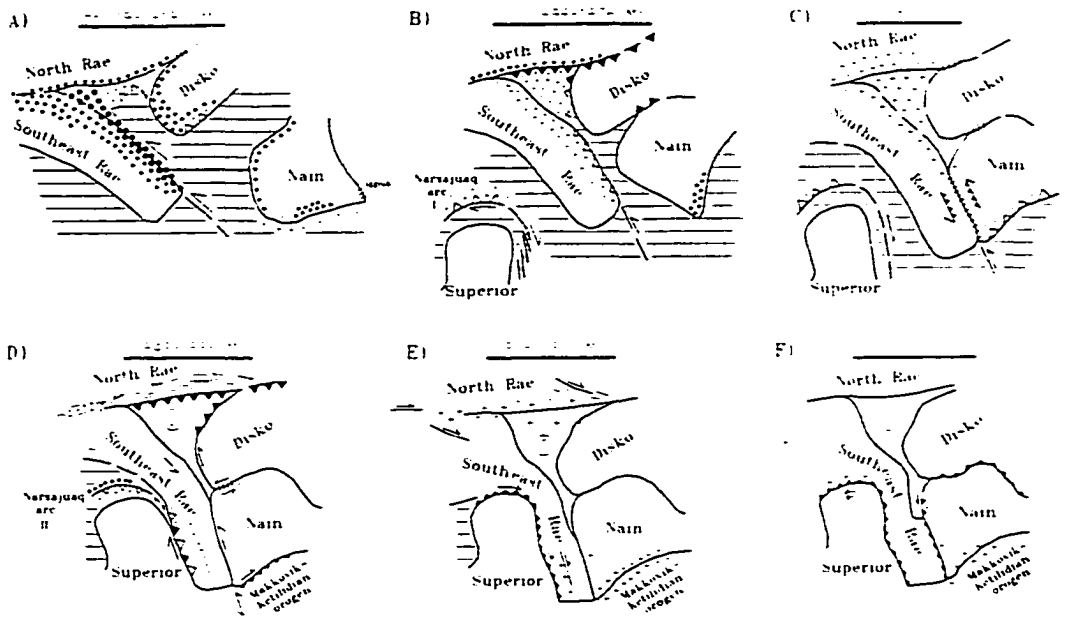


Figure 4. Schematic time-evolution diagrams fro the tectonic assembly of Northeast Laurentia between ca. 1.92-1.74 Ga from Van Kranendonk et al., 1993. Areas with horizontal lines are oceanic crust; small dots are platform sequences, turbiditic metasedimentary rocks and associated mafic rocks; large dots = Tasiuyak gneiss; x = sites of arc magmatism; + = late to post-tectonic granites; sutures = sites of collisional orogeny; white teeth = subduction zones; black teeth = thrusting; arrows indicate shear direction; circled C= Cumberland batholith complex.

The western margin of Nain Province collided with the eastern margin of the Southeastern Rae Province and intervening Tasiuyak gneisses at ca. 1860 Ma, initiating Torngat orogenesis. Subsequently, the western margin of the Southeastern Rae Province collided with the Superior Province along the New Quebec Orogen (1845-1829 Ma). Finally, northerly indentation of the Superior Province into the Northern Rae Province at ca. 1826-1758 Ma across the Ungava orogen marked the last of several indentations in northeastern Laurentia.

4.5 Models for development of the Torngat Orogen: different pre-collisional geometries

Since the publication of Van Kranendonk et al. (1993) model, a substantial amount of new geological information about the tectonic and chronologic development of the TO, particularly its northern-most segment, has become available. This information has been incorporated into several new models, each emphasizing different aspects of the evolution of the TO. These models can be divided into three catagories: (1) models suggesting westward subduction of Nain beneath Rae Province based on data and observations from the southern TO (e.g., Theriault and Ermanovics, 1997); (2) models proposing eastward subduction of Rae beneath Nain Province (e.g., Scott and Campbell, 1993; Scott, 1995a; Rivers et al., 1996; Ellis and Beaumont, 1996; Ellis et al., 1997) and (3) a model by Van Kranendonk and Wardle (1997), which employs either dual subduction or a flip in subduction polarity from east to west, in order to accomodate data from the two opposing models above. All models are described in detail below.

4.5.1 Westward subduction prior to collision - data from southern TO

Theriault and Ermanovics (1996) provide the most up-to-date synthesis of the geological features which characterize the southern TO from N58° to N57°30'. Early Proterozoic calc-alkaline metaigneous rocks occur within the eastern Rae Province (their "western metaplutonic suite") and within the Tasiuyak gneiss complex immediately east of the Rae Province ("eastern metaplutionic suite") and display "..abrupt tectonic contacts with all rocks of the orogen." (Theriault and Ermanovics, 1996, p. 19). Zircons from a sample from the eastern metaplutonic suite yield a series

of concordant U/Pb ages ranging from 1877 to 1822 Ma (Bertrand et al., 1993). The 1877 Ma age is interpreted as the emplacement age of the Early Proterozoic magmatic rocks throughout the area (Bertrand et al., 1993). The data are further interpreted to pin the Tasiuyak gneiss to the eastern edge of the Rae Province prior to 1877 Ma, and therefore prior to collision with the Nain Province at 1860 Ma (Bertrand et al., 1993). On the basis of these and additional observations, all from the southern TO, Theriault and Ermanovics (1996) propose that convergence between the Nain and Rae Province was accommodated by westward subduction beneath the Rae Province. They further argue that the westward decrease of initial εNd values for the Early Proterozoic metaigneous rocks in the southern TO is due to increased interaction of the arc magmas with basement gneisses from east to west (Fig. 5).

4.5.2 Eastward subduction prior to collision - new data and reinterpretations of existing data

Scott (1995a) provides the most current synthesis of data from both northern and southern TO in support of eastward dipping subduction beneath the Nain Province, as previously proposed by Van Kranendonk and Scott (1992), Scott and Campbell (1993) and Scott and Machado (1995). In this model, eastward dipping subduction occurred beneath the western margin of the Nain craton along the entire length of the TO between 1.91 and 1.86 Ga, followed by collision (1.86-1.84 Ga), and oblique uplift of the southern part of the Rae craton (1.80-1.71 Ga). This model is based on the speculation that the Burwell Domain is underlain by a western extension of the Nain Province (Scott, 1995b). The calc-alkaline rocks which intrude the Burwell Domain between 1.91 and 1.86 Ga (e.g., Scott, 1995a; Scott and Machado, 1995) must therefore represent a continental arc which developed in response to eastward subduction prior to collision with the Southeast Rae Province.

Scott (1995c) reinterprets the apparently conflicting data from the eastern and western metaplutonic suites in the southern TO (Bertrand et al., 1993; described in 5.1) and suggests that crustally derived magmas were emplaced at 1822 Ma and therefore are too young to be related to west-dipping subduction. The 1877 Ma zircons are interpreted by Scott (1995c) as inherited. The

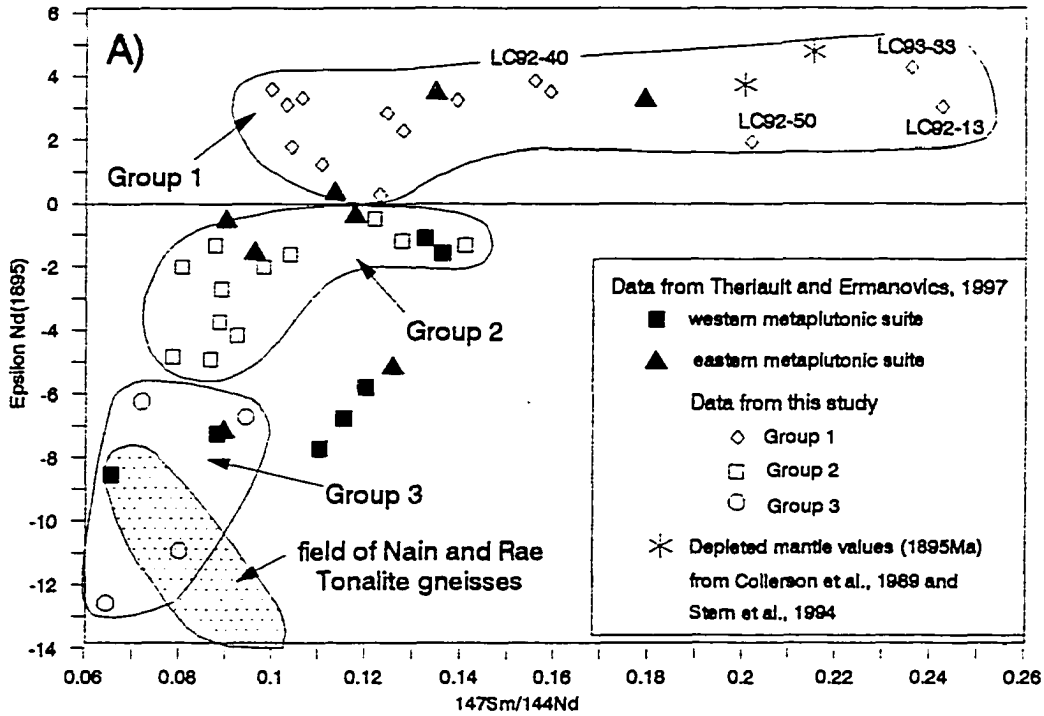


Figure 5. Sm/Nd isotope data for Early Proterozoic metaigneous rocks from northern and southern segments of the Torngat Orogen, Northern Labrador. $\varepsilon_{Nd\,(1895)}$ vs. $^{147}Sm/^{144}Nd$ diagram for samples from Group 1, 2 and 3 from the northern TO (Chapter 3 of this study) and for eastern and western metaplutonic suites from the southern TO (Theriault and Ermanovics, 1997). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994). Stippled field includes data from Nain and Rae Province Archean gneisses of the northern TO (Chapter 2). Data points for depleted mantle from Collerson et al. (1991) and Stern et al. (1994).

lack of ca. 1910-1864 Ma calc-alkaline plutonic rocks in the Nain craton south of the bend of the Abloviak shear zone (Fig. 6) is interpreted to be due to an along-strike difference in uplift and erosion, so that in the south, Early Proterozoic magmatic rocks that intruded the western margin of the Nain Province are largely eroded away. The resulting detritus from the ca. 1.91-1.86 Ga old rocks was subsequently deposited south of the present day exposures of the TO and later affected by Labradorian and Grenville orogenies.

4.5.3 Double subduction or flip in subduction-direction?

Van Kranendonk and Wardle (1997) developed two separate models to account for the occurrence of Early Proterozoic calc-alkalkine rocks on either side of the suture (e.g., within the Nain Province in the northern TO, and within the Rae Province and Tasiuyak gneiss complex in the southern TO). In model A, subduction was to the east between 1910 and 1885 Ma and the < 1885 Ma magmatic rocks occurring within the southwestern Burwell Domain are considered to represent syn-collisional magmas, generated during accretion of the Burwell arc onto the Nain Province. At 1885 Ma, subduction flips to the west, causing generation of the eastern and western metaplutonic suites in the southern TO (Theriault and Ermanovics, 1997). In model B, the younger (ca. 1865 Ma) magmatic rocks occurring in the northern TO are contemporaneous with the eastern and western metaplutonic suites in the southern TO, following a model of double subduction. In both models, large volumes of the 1910-1895 Ma calc-alkaline rocks were emplaced into the western margin of the Nain Province, thus forming a continental promontory. During subsequent Nain-Rae collision, the promontory was thrust eastward onto the Nain Province and tilted so that the western half of the promontory was buried to depths of ca. 35 km. Subsequent erosion resulted in the exposure of granulite facies assemblages in the western part and amphibolite facies assemblages in the eastern part of the Burwell domain (note: the present author's field and petrographic observations indicate no metamorphic east-west variation; see Chapter 3).

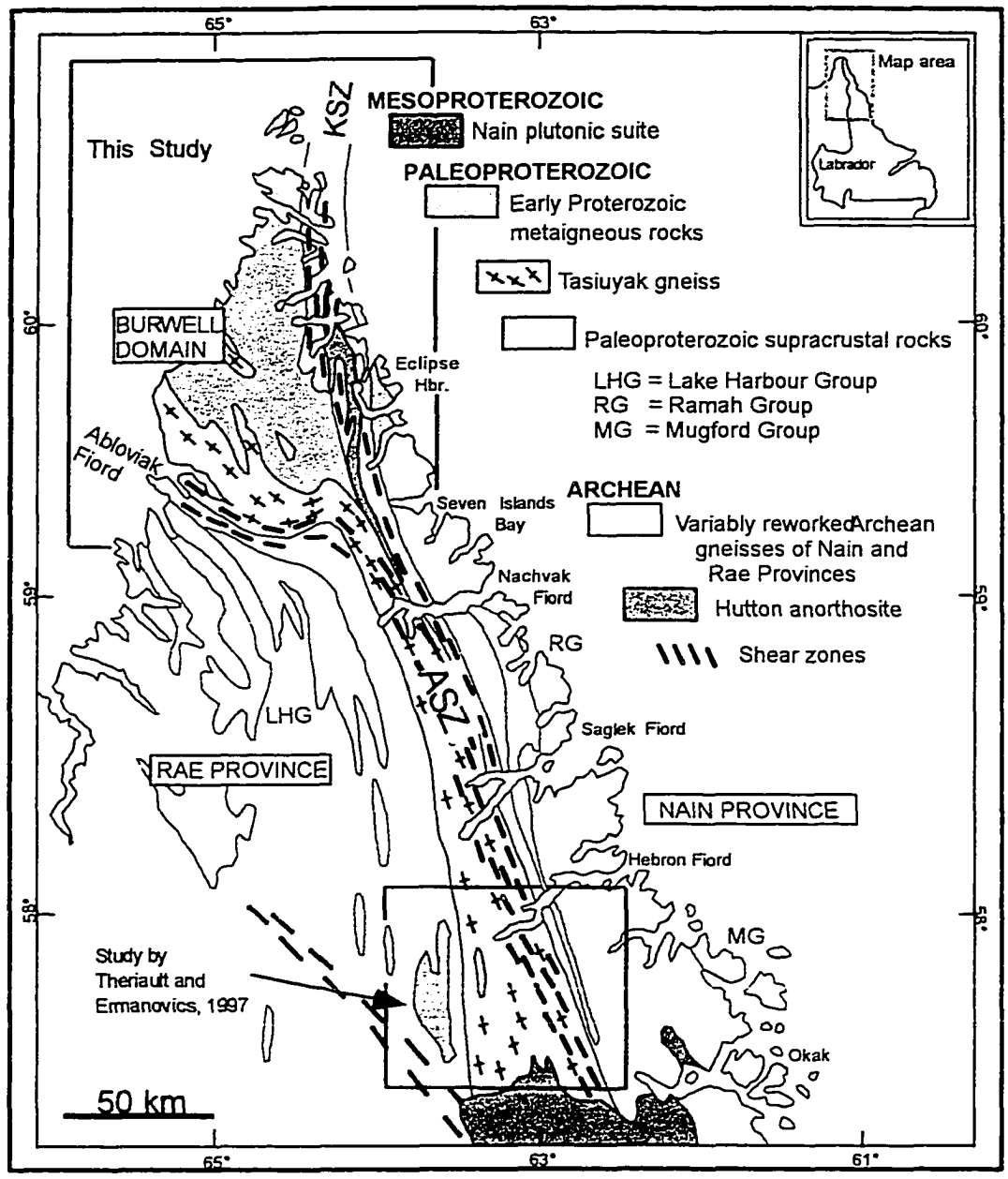


Figure 6. Simplified geological sketch map of Torngat Orogen, Northern Labrador, showing the main lithotectonic elements and geographic locations mentioned in the text (modified from Taylor, 1979; Korstgaard et al., 1987; Van Kranendonk & Ermanovics, 1990; Wardle et al., 1993; Wardle & Van Kranendonk, 1996). ASZ and KSZ are the Abloviak and Komaktorvik shear zones, respectively. Paleoproterozoic supracrustal successions are LHG = Lake Harbour Group, RG = Ramah Group, MG = Mugford group. Also shown are the location of Figure 2 ("This study") and the area described by Theriault & Ermanovics (1997).

4.6 Modifications to existing tectonic models: implications from geochemical and isotopic investigations

The new data and results presented in Chapter 2 and 3 require that several modifications be made to the above mentioned models for the assembly of Northeastern Laurentia. These new data and consequent modifications can be described in the following headings:

4.6.1 Origin of Early Proterozoic meta-igneous rocks: a magmatic arc built on the Nain margin - summary of data and interpretations

It was previously proposed by Scott and Campbell (1993), Rivers et al. (1996) and Van Kranendonk et al. (1994), that the Early Proterozoic magmas occurring in the Burwell domain were emplaced into Archean gneisses of the Nain Province. However, prior to the current study, there was no supporting geochemical or isotopic evidence for this. In the present study (Chapter 3), we have shown that the trace element character of the Early Proterozoic metaigneous rocks are essentially identical to that of Phanerozoic arc magmas, but are distinctly different from MORB. In addition, we have shown that Nd, Sr and Pb isotopic compositions of the Early Proterozoic metaigneous rocks range between those of a depleted mantle source and those of the Archean basement gneisses, indicating varying amounts of crustal contribution to a mantle-derived magma, as is expected in a continental arc setting. Field observations show that the Early Proterozoic metaigneous rocks intrude the Archean basement gneisses on either side of the Komaktorvik shear zone, but not to the west or south of the Abloviak shear zone. Based on this, and on the isotopic and geochemical data presented above, there is one model which successfully incorporates all features. In this, the Early Proterozoic metaigneous rocks formed as a result of eastward subduction beneath the western margin of the Nain Province, with the Abloviak shear zone marking the locus of the contact between the Nain and Rae cratons (i.e., the suture).

The lateral variation in initial Nd isotopic composition for the Early Proterozoic metaigneous rocks across the northern TO does not, however, show an increase in crustal contribution inland, and therefore does not conform to a simple Phanerozoic model for eastward

dipping subduction (Chapter 3). Instead, Early Proterozoic metaigneous rocks with initial Nd isotopic compositions similar to depleted mantle define a north-trending elongate array along the eastern margin of the Burwell domain. Peripherally around this core, Early Proterozoic metaigneous rocks have Nd isotopic compositions which suggest that the magmatic protoliths were derived from a depleted mantle source, but incorporated varying amounts of older felsic crust. In the western part of the Burwell domain, the Early Proterozoic rocks have Nd isotopic compositions similar to those of adjacent Archean basement gneisses. It was proposed in Chapter 3 that this distribution of Nd isotopic compositions reflects lateral variations in the composition of underlying crust into which the magmas were emplaced, such that along the eastern margin of the Burwell domain, the Archean basement gneisses were underlain by a thick pile of underplated basalt, probably emplaced during a period of incipient rifting at ca. 2.4 -2.2 Ga. The underplated basalt may have undergone partial melting to generate melts with mantle-like initial Nd isotopic compositions.

In support of the model proposed in Chapter 3, anorthosites and interlayered amphibolites occur only within the Komaktorvik shear zone, and may represent slices of the underplated basalt and its anorthositic differentiates, which were thrust into place before or during early phases of convergence between the Nain and Rae provinces. In addition, Early Proterozoic metaigneous rocks occur intrusively into the anorthosites and amphibolites, indicating that the anorthosites and amphibolites were already in place at the time of Early Proterozoic magmatism. These will be discussed further below.

4.6.2 The Burwell Domain: an Early Proterozoic magmatic complex built on Archean Nain Province.

The previous suggestion, that the Burwell domain is underlain by a western continuation of the Nain Province, was based mainly on similarity in field characteristics for Archean gneisses on either side of the Komaktorvik shear zone (Scott, 1995b; Van Kranendonk et al., 1994), and on a few late Archean zircon ages in the tonalitic gneisses west of the Komaktorvik shear zone, which

were similar to data obtained from Archean gneisses occurring east of the KSZ (Scott, 1995b: Connelly & Mengel, 1995).

In Chapter 2 of this study, we confirmed that Archean rocks occurring immediately west of the Komaktorvik shear zone are isotopically indistinguishable from Nain Province gneisses occurring to the east of it. It is not clear, however, how far this Nain-like crust extends towards the west. In the southwestern Burwell domain, Archean crustal blocks are isotopically dissimilar to the Nain Province, but are not unlike Archean gneisses of the adjacent Rae Province. Therefore we have proposed (Chapter 2), that the Archean crustal blocks in the southwestern part of the Burwell domain may represent crustal slices of Rae Province which were intercalated with the Burwell domain during Nain-Rae collision.

The conclusions from this work do not support earlier models which propose that the Burwell domain represents a foreign Archean craton (Hoffman, 1988, 1990), nor do they support current models suggesting that the western Burwell domain represents a Proterozoic magmatic arc which was accreted to the western margin of the Nain Province (Van Kranendonk and Wardle, 1997). In contrast, the initial Nd isotopic compositions of these western-most Early Proterozoic rocks indicate significant involvement with Archean felsic crust (Chapter 3).

4.6.3 Anorthosites and amphibolites: evidence for pre-collisional rifting?

Van Kranendonk and Wardle (1996; 1997) and Wardle and Van Kranendonk (1996) have previously suggested that rifting took place along the western margin of the Nain Province based on: (1) the occurrence of slivers of mafic supracrustal rocks with N-MORB-like REE-patterns within the Early Proterozoic metaplutonic rocks, and (2) the presence of the extensive belt of anorthosites occurring within the KSZ.

Data from the present study (Chapter 3) are in agreement with a period of rifting along the western margin of the Nain Province, and as described previously, we suggest that it took place prior to convergence and subduction, certainly prior to arc-magmatism. The location of Early Proterozoic metaigneous rocks with depleted, mantle-like initial Nd isotope compositions in a narrow, well-

constrained area along the eastern margin of the Burwell domain (Group 1 rocks), suggests the presence of a localized source. This feature is compatible with local basification of the crust by underplated basalt, presently represented at the surface by amphibolites and anorthosites. The subduction-related Early Proterozoic metaigneous rocks intrude the anorthosites and amphibolites as well as Nain Province gneisses, thus indicating that the anorthosites and amphibolites were already in place within the western Nain Province prior to arc magmatism. The occurrence of mafic dykes in the Archean gneisses on either side of the KSZ also lends support to the proposed rifting event. The presently available geochronological constraints suggest that dykes were emplaced in the period ca. 2.4-2.15 Ga (Cadman et al., 1993; Ermanovics, 1993; Connelly & Mengel, 1996; Cadman et al., 1997; Van Kranendonk and Wardle, 1997).

In order to justify a later period of rifting, specifically in a back-arc setting associated with eastward subduction beneath the Nain Province (as proposed by e.g., Van Kranendonk and Wardle, 1996), there should be evidence for multiple stages of arc magmatism (pre- and post-dating a period of back-arc rifting; e.g., Wilson, 1989). The arc rocks post-dating rifting (i.e., the youngest arc rocks) should show higher εNd values (similar to a depleted-mantle source) due to basaltic underplating and thinning of the Archean crust by back-arc rifting, prior to emplacement of these younger arc magmas into the rifted region. However, none of these relationships are observed. Instead, available data show that rocks with mantle-like Nd isotopic compositions (Group 1) represent some of the oldest arc rocks in the northern TO (ca. 1910 Ga; Scott, 1995a).

4.6.4 Tasiuyak gneiss: Passive margin sediments or an accretionary wedge?

The Tasiuyak gneiss represents a continuous belt of metasediments which can be traced along strike for >1300 km (e.g., Taylor, 1979; Hoffman, 1988). In northern Labrador it is largely coincident with the Abloviak shear zone, which marks the suture between the Nain and Rae provinces. The tectonic significance of the Tasiuyak gneiss is unclear, mainly because pervasive deformation along the ASZ has eradicated any primary depositional or structural contacts with the

adjacent Nain and Rae Province gneisses. Its significance or source characteristics can therefore only be inferred from secondary information such as isotopic composition and ages of detrital grains.

Nd isotopic compositions for the Tasiuyak gneiss are remarkably consistent from the northern to the southern TO, although in general, values in the north reflect a stronger contribution from Archean crustal material. In the southern TO, $\varepsilon Nd_{(1.9\ Ga)}$ ranges between -1.9 and -2.6 (Theriault and Ermanovics, 1997). In the northern TO, $\varepsilon Nd_{(1.895\ Ga)}$ ranges from -2.38 to -3.82 (Table 1). These values overlap with those of Group 2 Early Proterozoic metaigneous rocks occurring in the northern TO, ($\varepsilon Nd_{(1.895\ Ga)}$ values range between -1.1 and -5.1), but are significantly higher than those of Archean tonalitic gneisses ($\varepsilon Nd_{(1.895\ Ga)}$ values range between ca. -9 and -18). In an attempt to constrain the timing of deposition of the Tasiuyak gneiss and its provenance(s), Scott and Machado (1994), analyzed individual detrital zircon grains. The preliminary results indicate protolith ages from 2062 ± 2 Ma to 1940 ± 2 Ma, indicating that the depositional age of the protolith must be younger than 1940 Ma, but older than 1895 Ma (the age of a cross-cutting quartz diorite). This is consistent with the Nd isotopic data, which indicate that the Tasiuyak gneiss contains a significant component of Early Proterozoic detritus, however the detrital zircon ages (2062 ± 2 Ma, 2042 ± 4 Ma, 1940 ± 2 Ma), do not correspond to any known rocks in the region, including the Early Proterozoic metaigneous rocks (1910-1864 Ma). The detritus which formed the precursors of the Tasiuyak gneiss may have been derived from local sources no longer exposed, however at present, due to lack of evidence for such a source, we favor a distal source (i.e., not from the adjacent Nain or Rae cratons), implying that eroded material must have been transported significant distances and possibly deposited in an offshore trench adjacent to the Nain Province (Campbell, 1994; Scott & Machado, 1994; Scott, 1996; Theriault and Ermanovics, 1997). Scott and Ermanovics (pers. comm., 1996), propose deposition of the Tasiuyak gneiss precursors in an offshore trench setting analogous to the modern Sunda Trench (Moore et al., 1982), as this would account for the young age, remarkable strike length and compositional homogeneity of the Tasiuyak paragneiss.

TABLE 1 - Isotopic analyses of Tasiuyak gneiss, Torngat Orogen, Northern Labrador

	Sm (ppm)	Nd (ppm)	147Sm/ 144Nd	143Nd/ 144Nd(M)	143/144Nd (±2 sigma)	ENd (M)	T(DM) (Ga)
TASIUYAK GNEISSES							
Sample number							
LC92-27	9.42	59.5	0.0958	0.511258	7	-26.92	2.31
LC93-30	8.23	48.3	0.1030	0.511309	8	-25.92	2.40
LC92-14	4.87	30.6	0.0963	0.511191	8	-28.23	2.41

Analytical details are given in Appendix 1.

4.7 Early Proterozoic assembly of NE Laurentia: constraints from isotopic and geochemical studies in the northern Torngat Orogen - summary and conclusions

The new data and results provided by the isotopic and geochemical studies presented above (chapters 2 and 3) have allowed us to confirm aspects of the existing models for the evolution of the TO and have also provided new information about poorly- or unconstrained elements of these models. Below, we summarize the key data and discuss the implications for the Early Proterozoic evolution of NE Laurentia.

We have argued, based on the geographic distribution of isotope compositions in calc-alkaline arc-related plutons, that there was a period of basaltic underplating, possibly associated with rifting in northern Labrador. Physical evidence for this event is provided by the belt of amphibolites and anorthosites and although these rocks are tectonically emplaced into their present position, we suggest that their location in the eastern margin of the Nain Province broadly reflect the original position of the zone of underplated basalt and possible rifting. In addition, we see no evidence for back-arc rifting (i.e., syn- or post-dating arc-magmatism) in response to eastward subduction, as suggested by Van Kranendonk and Wardle (1996).

The timing of basalt underplating is poorly constrained, however, it pre-dates the oldest calc-alkaline intrusions (1910 Ma: Scott & Machado, 1995), which cut the anorthosites and amphibolites. The emplacement of mafic dykes in the region may be associated with basaltic underplating; this event is presently constrained at 2.4 - 2.15 Ga (see above). There is also evidence for a similar phase of Early Proterozoic basaltic magmatic activity elsewhere in the TO and its forelands:

(1) The Early Proterozoic Ramah Group, a supracrustal unit within the Nain Province which occurs between Nachvak and Saglek fiords (e.g., Morgan, 1975; Figure 5) contains a lower volcanic member and a gabbroic sill.

(2) The Early Proterozoic Mugford Group (e.g., Smyth, 1976; Figure 5) contains a large amount of basic volcanics in its upper part.

It should be noted that both the Ramah and Mugford groups were deposited on basement gneisses of the Nain Province and presently occur along the eastern margin and foreland of TO: analogous volcanic members are not known in any of the supracrustal series within or west of TO (i.e., Tasiuyak gneiss, Lake Harbour Group). Deposition of the Ramah and Mugford groups postdates emplacement of mafic dykes (Morgan, 1975; Smyth, 1976), hence these volcanic rocks may represent a relatively late stage of rifting and magmatism.

If related to rifting, this geographically widespread phase of basaltic magmatism may represent breakup of a Late Archean craton (as also speculated by Van Kranendonk and Wardle, 1997) and the range of ages may suggest that rifting was diachronous across NE Laurentia. Zones of rifting represent areas of thinned, i.e., weaker, continental crust, and therefore these are likely locations for later intra- and intercratonic deformation. The latter tectonic overprint would largely obliterate the rift-related structures and lithologies, which may explain the rare preservation and recognition of these features.

We suggest that (incipient) rifting, such as described above, was important for localizing the orogens in NE Laurentia. This speculation can be tested by an approach combining field observations and careful isotope geochemistry, similar to the study presented in Chapters 2 and 3.

The Burwell domain plays an important role in models for the Early Proterozoic evolution in northern Labrador and has therefore received special attention in this study. It is underlain by Archean basement gneisses throughout, and in the northern and eastern Burwell domain, these Archean basement gneisses represent a westward continuation of the Nain Province. Archean gneisses in the southwestern Burwell domain are, in contrast, isotopically dissimilar to the Nain Province, but resemble the adjacent Rae Province gneisses. This suggests that the southwestern Burwell domain represents a region of mixed gneisses from different provinces, presumably intercalated during collision between Nain and Rae Provinces. The Burwell domain is dominated by an Early Proterozoic continental arc which was constructed on the western margin of the Nain

Province, as previously proposed by Scott and Campbell (1993), Rivers et al. (1996) Van Kranendonk and Wardle (1997).

In light of the above, the Burwell domain does not represent an allocthonous (with respect to adjacent Nain and Rae Provinces) Archean craton as proposed by Hofmann (1988, 1990). Furthermore, it is clear that no portion of the Burwell domain represents an accreted Proterozoic island arc, as proposed by Van Kranendonk and Wardle (1997), hence their "orogenic boundary perturbation" is a primary feature of the Nain margin, rather than the result of an accreted arc.

The suture (defined as the boundary between the colliding Nain and Rae "continents") is usually placed coincident with the Tasiuyak gneiss. The recognition in this study, however, that Rae crust may occur as intercalated slices in the SW Burwell domain, would suggest that at the present level of exposure, the complex tectonic contact zone between Nain and Rae gneisses is located somewhere in the SW Burwell domain. Considering that this contact zone has been strongly modified by collisional tectonics, its precise location cannot be defined with the presently available data.

We suggest that in the Early Proterozoic, NE Laurentia developed through an early phase of rifting and localization of basaltic magmas, which locally penetrated the overlying crust (e.g., dykes, volcanics). During subsequent convergence between the cratons, calc-alkaline magmas were generated and intruded the continental margins, including the previously basified zones. During ensuing collision and transcurrent shearing, many of the pre-collisional features were obliterated. The models of Hoffman (1988, 1990) and Van Kranendonk et al. (1993) need to be modified to include this early phase of basaltic underplating/incipient rifting, and also to incorporate the Burwell domain as continental arc built on a continental margin rather than as a separate craton or accreted magmatic arc.

REFERENCES

Bertrand, J.M., Roddick, J.C., Van Kranendonk, M.J., Ermanovics, I., 1993. U-Pb geochronology of deformation and metamorphism across a central transect of the Early Proterozoic Torngat Orogen, North River map area, Labrador. Can. J. Earth Sci., 30: 1470-1489.

Bridgwater, D. Marker, M., Mengel. F., 1991. The eastern extension of the Early Proterozoic Torngat orogenic zone across the Atlantic. *In* ESCOOT Lithoprobe Report 27; 76-91.

Bridgwater, D., Austrheim, H., Hansen, B.T., Mengel, F., Pedersen, S., Winter, J., 1990. The Proterozoic Nagssugtoqidian mobile belt of southeast Greenland: a link between the eastern Canadian and Baltic Shields. Geosci. Can., 17:305-310.

Cadman, A. C., Heaman, L. M., Tarney, J., Wardle, R. J. and Krogh, T. E. (1993) U-Pb geochronology and geochemical variation within two Proterozoic mafic dyke swarms, Labrador. Can. J. Earth Sci., 30, 1490-1504.

Cadman, A. C., Tarney, J. and Hamilton, M. A. (1997) Petrogenetic relationship between Paleoproterozoic tholeiitic dykes and associated high-Mg noritic dykes, Labrador, Canada. Precambrian Res., 82, 63-84.

Campbell, L.M., 1994. Preliminary results from an Sm-Nd isotopic study of Proterozoic crustal formation and accretion in the Torngat Orogen, northern Labrador. Report of 1993 ECSOOT Transect Meeting. Edited by R.J. Wardle and J. Hall, Lithoprobe Report No. 34, p.100-107.

Chevé, S. R. and Machado, N. 1988. Reinvstigation of the Castignon Lake carbonatite complex, Labrador Trough, New Quebec. Geological Association of Canada - Mineralogical Association of Canada, Program with Abstracts, v. 13, p. A20.

Connelly, J. N., Mengel, F. C., 1995. Preliminary U-Pb data from Archean gneisses in a transect across the Komaktorvik Shear Zone, Northern Torngat Orogen. *In* ESCOOT Lithoprobe Report 45; 72-82.

Gaal, G., Gorbatschev, R., 1987. An outline of the Precambrian evolution of the Baltic Shield. Precambrian Res., 35:15-52.

Gorbatschev, R., Bogdanova., S., 1993. Frontiers in the Baltic Shield. Precambrian Res., 64: 3-21.

Grocott, J. and Pulvertaft, T. C. R. 1990. The Early Proterozoic Rinkian belt of central West Greenland, In: Lewry, J. F. and Stauffer, M. R. (editors) The Early Proterozoic Trans-Hudson Orogen of North America. Geological Association of Canada, Special Paper 37, pp. 443-463.

Henderson, J.R., 1985. Geology, Ekalugad Fiord-Home Bay, District of Franklin, Northwest Territories. Geological Survey of Canada, Map 1606A, scale 1:250,000.

Jackson, G.D., Hunt, P.A., Loveridge, W.D., and Parrish, R.R., 1990. Reconnaissance geochronology of Baffin Island, N.W.T. *In*: RadiogenicAge and Isotopic Studies, Rep.3, Geol. Surv. Can. Pap., 89-2:123-148.

Korstgaard, J. A., 1979a. Metamorphism of the Kangamiut dykes and the metamorphic and structural evolution of the southern Nagssugtoqidian boundary in the Itivdleq-Ikertoq region, West Greenland. Rapp. Gronlands geol. Unders. 89: 63-75.

Machado, N., Perrault, S., Hynes, A., 1988. Timing of continental collision in the northern Labrador Trough, Quebec: evidence from U-Pb geochronology. Geol. Assoc. Can., Progr. Abstr., 13:A76.

Mengel, F. C., 1984. Preliminary results of mapping in the Ramah Group and adjacent gneisses south of Saglek Fiord, northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 84-1: 21-29.

Mengel, F. C., 1985. Nain-Churchill Province boundary: a preliminary report on a cross-section through the Hudsonian front in the Saglek Fiord area, northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 85-1: 33-42.

Moore, G.F., Curray, J.R., Emmel, F.J., 1982. Sedimentation in the Sunda Trench and forearc region. *In:* J.K. Leggett Ed., Trench-forearc geology: sedimentation and tectonics on modern and ancient active plate margins. Geo. Soc. of London, Spec. Pap. 10:245-258.

Morgan, W. C. 1975. Geology of the Precambrian Ramah Group and basement rocks in the Nachvak Fiord-Saglek Fiord area, northern Labrador. Geological Survey of Canada, Paper 74-54, 42pp.

Park, R.G., 1997. Early Precambrian plate tectonics. S. Afr. J. Geol., 100/1: 23-35.

Park, R.G., 1994. Early Proterozoic tectonic overview of the northern British Isles and neighbouring terrains in Laurentia and Baltica. Precambrian Res., 68:65-79.

Peltzer, G., Tapponier, P., 1988. Formation and evolution of strike-slip faults, rifts and basins during the India-Eurasia collision: an experimental approach. J. Geophys. Res., 77:4432-4460.

Perriault, S., Hynes., 1990. Tectonic evoution of the Kuujjuaq terrane, New Quebec Orogen. Geosci. Can., 17: 238-241.

Perriault, S., Hynes, A., Machado, N., 1988. Timing in the tectonic evolution of the NE segment of the Labrador trough, Kuujjuak, northern Quebec. Geol. Assoc. Can., Progr. Abstr., 13:A97.

Pidgeon, R.T., Howie, R.A., 1975. U-Pb age of zircon frm a charnockitic granulite from Pangnirtung on the east coast of Baffin Island. Can. J. Earth Sci., 12:1046-1047.

Ryan, B., Martineau, Y., Bridgwater, D., Schiotte, L., Lewry, J., 1983. The Archaean-Proterozoic boundary in the Saglek fiord area, Labrador: Report 1. *In* Current Research, part A, Geol. Surv. Canada Paper 83-1A: 297-304.

Ryan, A. B., Martineau, Y., Korstgaard, J., Lee, D., 1984. The Archaean-Proterozoic boundary in northern Labrador, Report 2. *In* Current Research, part A, Geol. Surv. Canada Paper 84-1A: 545-551.

Scott, D.J., 1995a. U-Pb geochronology of a Paleoproterozoic continental magmatic arc on the western margin of the Archean Nain craton, northern Labrador, Canada. Can. J. Earth Sci., 32: 1870-1882.

Scott, D.J., 1995b. U-Pb geochronology of the Nain craton on the eastern margin of the Torngat Orogen, Labrador. Can. J. Earth Sci., 32: 1859-1869.

Scott, D. J. (1995c) U-Pb geochronology of the northeastern Rae Province. Part 1: A preliminary report on the Lac Lomier Complex, Northern Labrador. In: Wardle, R. J. & Hall, J. (compilers). Eastern Canadian Onshore-Offshore Transect (ECSOOT), Transect Meeting, École Polytechnique, Montréal, Québec, Report 45: 122-128.

Scott, D. J. and Machaco, N. (1994) U-Pb geochronology of the northern Torngat Orogen: results from work in 1993. In: Wardle, R. J. & Hall, J. (compilers). Eastern Canadian Onshore-Offshore Transect (ECSOOT), Transect Meeting, The Université du Québec à Montréal, Report 36: 141-155.

Scott, D. J., Campbell, L. M., 1993. Evolution of the Paleoproterozoic Torngat Orogen, Labrador, Canada: Recent advances using U-Pb geochronology and Nd isotope systematics. *In* Geol. Soc. Am. Program with Abstracts, vol. 25, A-237.

Scott, D. J., Helmstaedt, H. and Bickle, M. J. 1992. Purtuniq ophiolite, Cape Smith belt, northern Quebec, Canada: a reconstructed section of Early Proterozoic oceanic crust. Geology, 20: 173-176.

Scott, D. J., St-Onge, M. R., Lucas, S. B. and Helmstaedt, H. 1991. Geology and chemistry of the early Proterozoic Purtuniq Ophiolite, Cape Smith Belt, northern Quebec, Canada. In: Peters, Tj., Nicolas, A. and Coleman, R. J. (editors), Ophiolite genesis and evolution of the oceanic lithosphere. Dordrecht, Klüwer, Paleoproterozoic. 817-849.

Smyth, W. R. 1976. Geology of the Mugford Group, northern Labrador. In Report of Activities for 1976. Newfoundland Department of Mines and Energy, Report 77-1, 72-79.

St-Onge, M.R., Lucas, S.B., 1990. Evolution of the Cape Smith Belt: Early Proterozoic Continental Underthrusting, Ophiolite Obduction and Thick-skinned Folding *In*: Lewry, J.F., and Stauffer, M.R., eds., The Early Proterozoic Trans-Hudson Orogen of North America: Geol. Assoc. of Can., Spec. Pap. 37:313-351.

Taylor, F.C., 1979. Reconnaissance geology of a part of the Precambrian Shield in northeastern Quebec, northern Labrador and Northwestern Territories. Geologica Survey of Canada, Memoir 393: 99pp.

Theriault, R.J., Ermanovics, I., 1996. Sm-Nd isotopic and geochemical characterization of the Paleoproterozoic Torngat orogen, Labrador, Canada. Precambrian Res., 81: 15-35.

Van Kranendonk, M. J. and Wardle, R. J., 1996. Burwell domain of the Paleoproterozoic Torngat Orogen: Tilted cross-section of a magmatic arc caught between a rock and a hard place. *In* Precambrian Crustal Evolution in the North Atlantic Region. *Ed.* T. S. Brewer. Spec. Publ. Geol. Soc. Lond. 112: 91-116.

Van Kranendonk, M.J., Wardle, R.J., (in press). Crustal scale flexural slip folding during late tectonic amplification of an orogenic boundary perturbation in the Paleoproterozoic Torngat Orogen, NE Canada. Can. J. Earth Sci. (1997)

Van Kranendonk, M.J., Wardle, R.J., Mengel, F., Campbell, L.M., Reid, L., 1994. New results and summary of the Archean and Paleoproterozoic geology of the Burwell domain, northern Torngat Orogen, Labrador, Quebec, and Northwest Territories. Current Research, Geological Survey of Canada, 1994-C: 321-332.

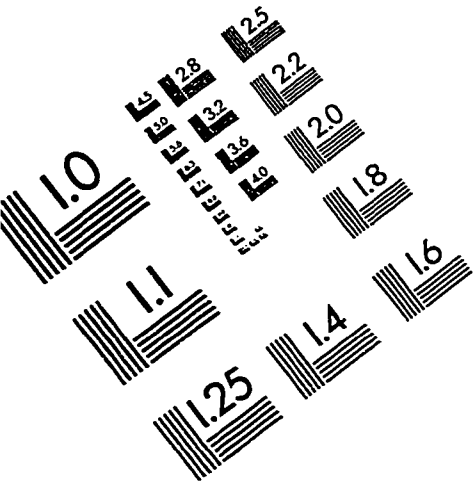
Van Kranendonk, M. J., Scott, D. J., 1992. Preliminary report on the geology and structural evolution of the Komaktorvik Zone of the Early Proterozoic Torngat Orogen, Eclipse Harbour area, northern Labrador. *In* Current Research, Part C. Geological Survey of Canada, Paper 92-1C: 59-68.

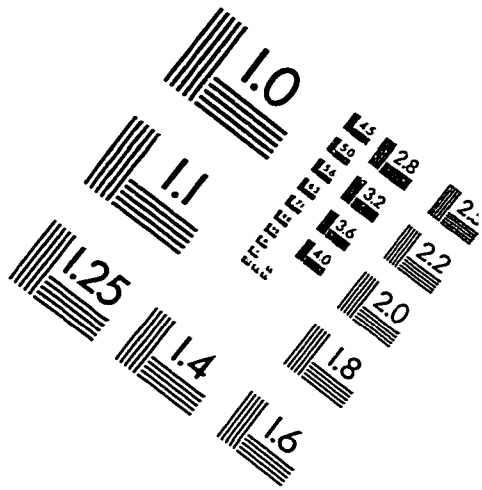
Wardle, R.J., Van Kranendonk, M.J., 1996. The eastern Churchill Province of Labrador-Quebec, Canada: Orogenic development as a consequence of oblique collision and indentation. *In*: Precambrian Crustal Evolution in the North Atlantic Regions. Eds. T.S. Brewer, B.P. Atkin. Spec. Pub of the Geol. Soc. of London.

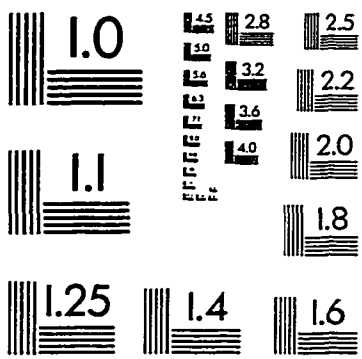
Wardle, R. J., 1983. Nain-Churchill province cross-section, Nachvak fiord, Northern Labrador. *In* Current Research, Newfoundland Department of Mines and Energy, Report 83-1: 68-89.

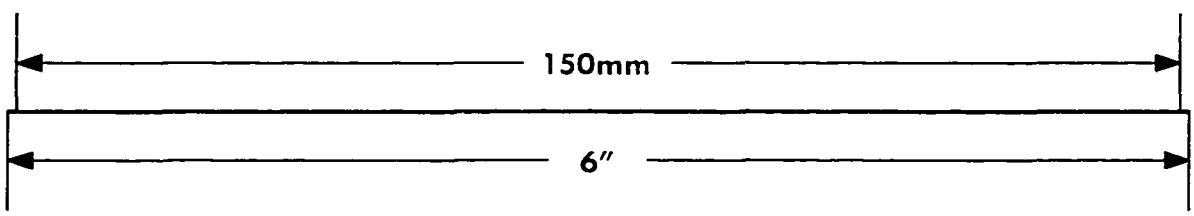
Wardle, R. J., 1984 Nain-Churchill province cross-section, Riviere Baudancourt-Nachvak Lake. *In* Current Research, Newfoundland Department of Mines and Energy, Report 84-11: 1-11.

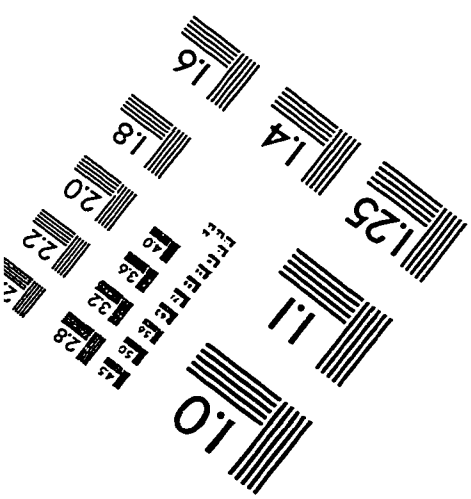
IMAGE EVALUATION TEST TARGET (QA-3)

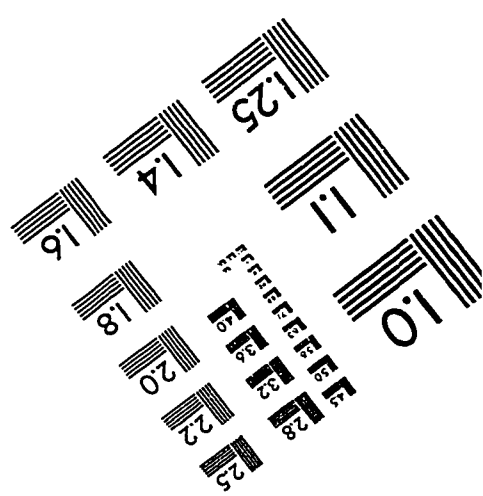












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

CHARACTERIZATION OF STRUCTURAL DOMAINS OF Rap1p AND THEIR INVOLVEMENT IN HIGHER ORDER TELOMERIC STRUCTURE IN YEAST

by

ANN CASSIDY-STONE

B.S., University of San Francisco, 1969

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

1998

This thesis entitled:
Characterization of Structural Domains of
Rap1p and Their Involvement in Higher
Order Telomeric Structure in Yeast
written by Ann Cassidy-Stone
has been approved for the Department of Chemistry and
Biochemistry

Steve Schultz

Olke Uhlenbeck

Date 3/12/98

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Cassidy-Stone, Ann (Ph.D. Chemistry and Biochemistry)
Characterization of Structural Domains of Rap1p and Their Involvement in Higher Order Telomeric Structure in Yeast
Thesis directed by Associate Professor Steve C. Schultz

Repressor Activator Protein 1 (Rap1p) of the yeast *S. cerevisiae* is an essential sequence-specific double-stranded DNA binding protein which is a transcriptional activator, a transcriptional silencer, and a telomere binding protein. Rap1p participates in formation of a large nucleoprotein structure at telomeres and at the silent mating-type loci. Our studies have explored the structure and DNA binding properties of Rap1p to better understand how Rap1p participates in assembling and maintaining higher order chromatin structures in yeast.

We developed methods for producing large quantities of highly purified Rap1p and subjected this material to limited proteolysis to define three structural domains: an N-terminal domain (NTD), a DNA binding domain (DBD) and a C-terminal domain. The N-terminal domain and DNA binding domain, as well as the NTD plus DBD and DBD plus C-terminal domain were subcloned, overexpressed and purified for use in crystallization and biochemical characterizations.

Our crystallization efforts focused on Rap1p and its structural domains complexed with DNA. Crystals of Rap1p

DNA binding domain plus C-terminal domain complexed with a 20 base pair DNA were obtained. These crystals diffract to 4.2A resolution and are cubic, space group I432. Also obtained were crystals of the Rap1p DNA binding domain complexed with a 19 base pair DNA. These crystals diffracted to 3.2 A resolution and are orthorhombic, space group P222. Crystals of the N-terminal domain alone also were obtained, but the crystals were too small for x-ray analysis.

Although it is generally believed that Rap1p exists and functions as a monomer, we found, using gel filtration analysis and Ferguson analysis of electrophoretic mobility in native gels, that Rap1p forms very stable dimers and tetramers in solution. Analysis of the structural domains of Rap1p demonstrates that the N-terminal domain is necessary and sufficient for Rap1p dimerization and tetramerization. Previously, no function had been ascribed to the N-terminal domain and, indeed, the NTD was believed to be dispensable for function.

Rap1p dimerization and tetramerization has important implications for formation of higher order structures at yeast telomeres. Although previous studies had indicated that deleting the N-terminus had no phenotypic consequences, we wished to examine this more carefully. We constructed a new Rap1p deletion strain and then examined the effect of a small N-terminal deletion, Rap1Δ10p, and complete deletion of the N-terminal domain, rap1DBDTELp, on telomeric silencing, mating efficiency and localization of Rap1p *in vivo*. These N-terminal deletions significantly reduce silencing of a telomere-

proximal gene, enhance mating efficiency, and give rise to a more diffuse staining of telomeric Rap1p. These findings suggest that dimerization and tetramerization of Rap1p as mediated by the N-terminal domain may indeed be functionally important in yeast.

Dedication

This thesis is dedicated to my husband, Walt Stone, whose love and encouragement kept me going through the adventure of graduate school, and to my brother, Frank Cassidy, who is a model of gracious determination in the face of adversity.

Acknowledgments

My advisor, Dr. Steve Schultz, has been a supportive and enthusiastic mentor. He has created a lab environment that is conducive to scientific inquiry and cooperativity among lab members. This atmosphere has been ideal for my growth as a scientist. Steve has also been very generous with his time. I consider Steve a friend as well as an advisor.

Dr. Lorraine Pillus suggested the Rap1p project to me and gave me the clone that allowed me to begin working on it. Throughout my work, Lorraine has offered me a great deal of encouragement and she generously allowed me to do the experiments described in chapter six in her laboratory. Lorraine's support has been very important to me.

I am grateful to the members of the Schultz lab for their enthusiasm about science and life in general. Martin Horvath wrote the computer program I used for analysis of the two-site cooperativity data. Eric Rosenblum ran gels for the binding studies described in chapters two and three. Martin Horvath, Olve Peersen, Scott Classen, and Andy Berglund provided invaluable assistance with computer graphics. My classmates Jodi Ryter and Jeff Hansen were an especially important part of my support system.

I am also grateful to members of the Pillus lab, who were very generous in helping me with the yeast genetics experiments described in chapter six. Sandi Jacobson helped me construct the *rap1* knockout strains, Elisa Stone assisted me with the immunolocalization experiments, and Joanna Lowell did the Southern blot to examine telomere length. Astrid Clark generously shared her bench space with me and taught me not only to dissect tetrads but also to have fun doing it.

My colleagues in the Schultz and Pillus labs have been wonderful companions on the journey.

CONTENTS

CHAPTER

I INTRODUCTION....1

II LARGE SCALE PRODUCTION OF Rap1p IN *E. coli* AND CHARACTERIZATION OF ITS DNA BINDING PROPERTIES AND DOMAIN STRUCTURE

Introduction....18

Materials and Methods

Expression and Purification of Rap1p....22

Limited Proteolysis....24

Expression and Purification of Rap1p Structural Domains....26

Binding Studies....30

Results

Rap1p Purification....32

Identification of Structural Domains....32

DNA binding properties....35

Discussion....46

III COOPERATIVE BINDING OF Rap1p TO ADJACENT DNA BINDING SITES DEPENDS ON SPACING BETWEEN SITES

Introduction....51

Materials and Methods

Binding Experiments....54

Data Analysis....56

Results....58

Discussion....64

IV CRYSTALLIZATION

Introduction....69

Materials and Methods

Crystallization trials....72

Rap1 DBDTEL complexed with DNA....72

Rap1 DBD complexed with DNA....75

RAP1 N-terminal domain....75

Diffraction analysis....76

Results

Rap1 DBDTEL complexed with DNA 8....76

Rap1 DBD complexed with DNA 9....78

Rap1 N-Terminal Domain....78

Discussion....79

V Rap1p N-TERMINAL DOMAIN MEDIATES OLIGOMERIZATION

Introduction....83

Materials and Methods

Gel Filtration Column Experiments....86

Ferguson Gels....87

Construction of plasmids encoding RAP1 N-terminal modifications.....89

Purification of Rap1p N-terminal Modifications.....90

Proteolysis of Rap1Δ10.....91

Rap1Δ10 Binding Experiments.....92

Results

Gel Fitration Analysis.....93

Ferguson Analysis.....98

Rap1Δ10, Rap1Δ15, and Rap1Nlys7: DNA Binding and Proteolysis.....102

Discussion.....105

VI DELETION OF Rap1p N-TERMINAL DOMAIN AFFECTS TELOMERE POSITION EFFECT, Rap1p IMMUNOLOCALIZATION, MATING EFFICIENCY, AND TELOMERE LENGTH

Introduction.....119

Materials and Methods

Plasmid construction.....123

Strain Construction.....126

Dilution assay.....130

Quantitative Mating Assay.....130

Assessment of *ADE2* transcription from telomere-proximal locus.....131

Immunofluorescence staining.....131

Determination of telomere length.....134

Results

RAP1 knockout.....135

rap1Δ10 and *rap1DBDTEL* viability, temperature sensitivity.....136

Telomere position effect.....136

Western Blot.....141

Immunofluorescence.....141

Telomere length.....143

Mating.....143

Discussion.....147

BIBLIOGRAPHY.....160

TABLES

Table

2.1 Summary of Determination of Rap1p Structural Domains....38

2.2 Dissociation Constants for Rap1p, Rap1 DBD, Rap1 DBDTEL, RAP1 NTD+DBD, and Rap1 NTD....46

3.1 Hill Coefficients and Dissociation Constants for Rap1p Binding to DNAs Containing Two Adjacent Binding Sites....61

3.2 Hill Coefficients and Dissociation Constants for Rap1 DBDTEL Binding to DNAs Containing Two Adjacent Binding Sites....63

3.3 Hill Coefficients and Dissociation Constants for Rap1 DBD Binding to DNAs Containing Two Adjacent Binding Sites....65

4.1 DNAs Used in Crystallization Trials....73

5.1 Elution Volumes From the Size Exclusion Column....96

5.2 Negative Slopes Obtained From Ferguson Analysis....101

5.3 Mixing Experiments....110

6.1 Growth on 5-FOA....139

6.2 Expression of a Telomere-Proximal *ADE2* Reporter Gene....140

6.3 Number of Rap1p Foci Visible per Cell Using Immunofluorescence Microscopy....144

6.4 Mating Efficiency....146

Figures

Figure

1-1 Four Models of Yeast Telomeres....13

2-1 Rap1p Consensus Sequences....20

2-2 Rap1p Purification....33

2-3 Limited Proteolysis of Rap1p Reveals Three Stable Structural Domains....34

2-4 N-terminal Sequences of Proteolytic Fragments....36

2-5 Western Blot of Limited Proteolysis Gel....37

2-6 Points of Cleavage by α-chymotrypsin and Trypsin....39

2-7 Rap1p Structural Domains....40

2-8 Rap1p Binding Gels....41

2-9 Rap1p Binding Curves....43

2-10 Effect of Dilution Buffer Without BSA On Rap1 DBD Binding....44

2-11 Rap1 DBD Binding Equilibrium....44

3-1 Rap1p Bound to RBS+8....59

4-1 Crystals of Rap1 DBDTEL Bound to a 20 Base Pair DNA....77

4-2 Crystals of Rap1 DBD Bound to a 19 Base Pair DNA....77

5-1 Chromatograph of Rap1p Elution From Gel Filtration Column....94

5-2 Gel Filtration Column Analysis of Rap1p....95

5-3 Chromatograph of Rap1Δ10 Elution From Gel Filtration Column....97

5-4 Ferguson Analysis of Rap1p....99

5-5 Sample Ferguson Gels....100

5-6 Rap1p and Rap1Δ10p Gel Shift Experiment....103

5-7 Limited Proteolysis of Rap1Δ10p....104

5-8 Rap1p N-terminal Modifications....112

5-9 Model of Rap1p Assembly of Telosome....117

6-1 Growth of *RAP1* and Mutant Strains on 5-FOA....138

6-2 Images of Intact Haploid Yeast Spheroplasts....142

6-3 Effects of *rap1DBDTEL* and *rap1Δ10* on Telomere Length....145

GLOSSARY OF TERMS

A - adenosine

BSA - Bovine Serum Albumin

C - cytosine

ds - double-stranded

DNA - deoxyribonucleic acid

DTT - dithiothreitol

EDTA - ethylenediamine tetraacetatic acid

FPLC - Fast protein liquid chromatography

g- gram

Hepes - 4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid

HPLC - High performance liquid chromatography

IPTG - isopropyl β-D-thiogalactopyranoside

Kbp - 1000 base pairs

KOH - potassium hydroxide

µg - microgram

mg - milligram

ml - milliliter

mm - millimeter

L - liter

PAGE - polyacrylamide gel electrophoresis

PBS - Phosphate Buffered Saline (.14 M sodium chloride, 3 mM potassium chloride, 1.8 mM potassium phosphate

(pH 7.4)

PEG - polyethylene glycol

SDS - sodium dodecyl sulfate

TEMED - N, N, N', N' tetramethyl-ethylenediamine

Tris - Tris (Hydroxymethyl) aminomethane

TBE - .09 M Tris-borate, 2 mM EDTA

Chapter I

Introduction

All organisms continuously face critical decisions on which genes to express and which to maintain in a repressed state. The rich variety and complexity of mechanisms cells utilize in regulating gene expression at the level of transcription can be grouped into three categories: direct effects on RNA polymerase activities, regulation of RNA polymerase interactions at specific promoters, and establishment of chromatin structures that determine access of factors necessary for transcription. The regulation of gene expression by altered chromatin structures is a universal phenomenon in eukaryotic cells and is responsible for implementation of the development program in multicellular organisms, for position effect variegation in flies, and for silencing at the *HM* loci and at telomeres in yeast (Gotta and Gasser, 1996). Genetic and biochemical studies of yeast telomeres and silent mating type loci have identified a common set of proteins required for silencing in yeast. These include Repressor Activator Protein 1 (Rap1p), Silent Information

Regulator Proteins (Sir2p, Sir3p, Sir4p), Nat1p/Ard1p (subunits of an N-terminal acetyl transferase), and histones H3 and H4 (Aparicio et al., 1991; Hecht et al., 1995; Laurenson and Rine, 1992). Identification of these components is leading to a molecular understanding of the higher order chromatin structures involved in *HML*, *HMR* and telomeric silencing in yeast. Models have been proposed in which Rap1p binds to DNA and recruits the Sir proteins which bind the N-terminal histone tails and set up a condensed chromatin structure. The research described here has focused on a central component of this condensed chromatin structure, Rap1p, which directly interacts with DNA and directs assembly of other factors involved in silencing.

Rap1p is a multifunctional protein that was originally isolated four times for four different functions: Translational Upstream Factor 1 (Tuf1p) (Huet et al., 1985), Telomere Binding protein A (Tba) (Berman et al., 1986), Silencer Binding Factor E (Sbf-e) (Shore and Nasmyth, 1987), and General Regulatory Factor 1 (Grf1p) (Buchman et al., 1988). Depending entirely on the context of its binding site, Rap1p can act as a

transcriptional activator or a transcriptional repressor (Shore and Nasmyth, 1987). Switching a gene normally activated by Rap1p for one normally silenced, as well as swapping cis-acting sequences (the silencer for a UAS) does not affect Rap1p activity at that location. It is the context in which Rap1p binds, not the gene upon which it is acting, that determines its function (Brand et al., 1987; Buchman et al., 1988; Giesman et al., 1991; Shore, 1994).

Rap1p activates transcription of many ribosomal protein and glycolytic enzyme genes in yeast (Kraakman et al., 1989; Vignais et al., 1990). The promoters of at least 75% of ribosomal protein genes contain Rap1p binding sites in their upstream activating sequence (UAS) known as the Ribosomal Protein Gene (RPG) box (Warner, 1989). Transcriptional activation of several glycolytic enzyme and ribosomal protein genes, including *ADH1, PGK1, PYK1, TEF1* and *TEF2*, involves Rap1p recruitment of Gcr1p, a *trans*-activator which is required for full activation of these genes (Drazinic et al., 1996; Tornow et al., 1993). In the case of transcriptional activation of the *MAT*α locus, however, Rap1p appears to be the only

transactivator that binds to the upstream activating sequence and this UAS is sufficient to activate transcription of a test gene containing only a TATA box (Brand et al., 1985; Giesman et al., 1991).

Rap1p is also involved in transcriptional repression at the silent mating-type loci. Rap1p, together with AbfIp and the origin recognition complex (ORC), binds to the cis-acting silencer elements at *HMR* and *HML*, contributing to the repression of the mating type genes at these loci (Laurenson and Rine, 1992). *Saccharomyces cerevisiae* chromosome III contains three loci which encode mating information: the two silent *HM* loci and *MAT* from which cell-type specific genes are expressed. *S. cerevisiae* can switch between **a** and α cell types, using mating-type information stored at *HML* and *HMR* in this mating-type interconversion. Mating-type interconversion is initiated when the HO endonuclease produces a double-stranded cleavage at *MAT* and is completed when repair of the break leads to a duplicative transposition of information from *HML* or *HMR* to *MAT* (Herskowitz, 1988). Rap1p binding plays a role in maintaining *HML* and *HMR* in the transcriptionally silent state necessary for the cell

to display normal mating behavior.

Rap1p is also a telomere binding protein which contributes to telomere length regulation as well as to the silencing of genes proximal to telomeres which is referred to as Telomere Position Effect (Gottschling et al., 1990; Kyrion et al., 1992; Liu et al., 1994). Rap1p binding sites occur in tandem repeats, reportedly as often as every 18 base pairs (Gilson et al., 1993) in the double-stranded $(C_{1-3}A)_n$ tract of yeast telomeres (Buchman et al., 1988; Buchman et al., 1988; Krauskopf and Blackburn, 1996; Longtine et al., 1989). Through the use of immunofluorescence microscopy and fluorescence *in situ* hybridization, it has been demonstrated that Rap1p is primarily localized to telomeres which are observed in distinct foci at the nuclear periphery (Cockell et al., 1995; Gotta et al., 1996; Klein et al., 1992; Moretti et al., 1994; Palladino et al., 1993). This suggests that Rap1p may play a role in forming a higher order structure at yeast telomeres and may be involved in telomere organization and localization.

Transcriptional silencing at telomeres and at the silent mating type loci involves many of the same

proteins. These are the Silent Information Regulator Proteins (Sir2p, Sir3p, Sir4p), Nat1p/Ard1p (subunits of an N-terminal acetyl transferase), and histones H3 and H4 (Aparicio et al., 1991; Hecht et al., 1995; Laurenson and Rine, 1992). In addition, silencing at the silent mating-type loci and telomeres appears to be linked to a highly condensed chromatin structure. Chromatin at *HML* and *HMR* is less accessible to nucleases including DNaseI, MnaseI (Nasmyth, 1982), and yeast HO endonuclease cleavage than chromatin from the *MAT*α locus. Likewise, Loo *et al* have reported a Sir-dependent resistance of chromatin at the *HM* loci to ten restriction endonucleases (Loo and Rine, 1994). Telomeric DNA also appears to be highly condensed in that it is less accessible to ectopically expressed *E. coli dam* methyl transferase than most of the yeast genome (Gottschling, 1992).

A higher order, non-nucleosomal structure that comprises yeast telomeres has been referred to by Zakian and colleagues as the telosome (Wright et al., 1992). This structure can be cleaved from chromosomal DNA at a DNaseI hypersensitive site to yield soluble particles which can be immunoprecipitated using antibody to Rap1p.

The gel mobility shift characteristic of soluble telosomes can be recapitulated by complexes formed *in vitro* between yeast telomeric DNA and purified Rap1p, suggesting that Rap1p is a major structural protein of the telosome (Wright and Zakian, 1995). Rap1p is known to bend DNA (Muller et al., 1994; Vignais et al., 1989), and there is evidence that it mediates loop formation *in vitro* at *HML* (Hofmann et al., 1989). One can easily imagine that features such as DNA bending and loop formation might be important in assembling higher order nucleoprotein complexes that constitute larger cell structures such as the telosome.

Biochemical, genetic, and cell biological studies have provided a wealth of information about the proteins that act at telomeres as well as at *HML* and *HMR*. Immunofluorescent staining reveals that Sir2p ,Sir3p, Sir4p, and Rap1p colocalize with telomeres in discrete, brightly stained foci. Although the diploid nucleus contains 64 telomeres, only ~8 foci are observed and these appear at the nuclear periphery. In addition to its telomeric localization, Sir2p is found in a subdomain of the nucleolus. In yeast strains containing a *sir3*

deletion or a *sir4* deletion, Rap1p colocalizes with yeast telomeres but is also seen as diffusely staining the entire nucleus. In these strains, Sir2p completely loses its telomeric localization, but retains the nucleolar localization. In a *sir3Δ/sir3Δ* strain, Sir4p loses its telomeric localization, whereas in a *sir4Δ/sir4Δ* strain, Sir3p loses its telomeric localization and is relocated to the nucleolus in a Sir2p-dependent manner. Telomeres in *sir3Δ/sir3Δ* or *sir4Δ/sir4Δ* strains still appear clustered, but no longer have the tightly defined foci seen in wild-type cells. In these mutant strains, the rate of chromosome loss is increased 2-5 fold over wild-type (Cockell et al., 1995; Gotta et al., 1996; Klein et al., 1992; Moretti et al., 1994; Palladino et al., 1993). In a *sir2Δ/sir2Δ* strain, Rap1p foci become less discrete and Sir3p and Sir4p staining appears diffuse, suggesting that Sir2p is an integral component of telomeric chromatin (Gotta et al., 1997). Overexpression of Sir4p causes pronounced Rap1p and Sir3p delocalization (Cockell et al., 1995). In histone H3(Δ4-30) or H4(Δ4-28) mutant strains, the Rap1p staining pattern is less discrete and is delocalized relative to wild-type cells. In these

same strains, Sir3p staining has lost all punctate character and perinuclear localization and become diffuse throughout the nucleus. These results suggest that telomere-telomere interactions as well as specific subnuclear telomeric localization are dependent on Rap1p, Sir2p, Sir3p, Sir4p, histone H3 and histone H4, and that these proteins are all components of yeast telomeres.

Silencing at telomeres and at *HMR* and *HML* has been shown to be dependent on Rap1p, Sir3p, Sir4p, and histones H3 and H4. The positively charged tails of histones H3 and H4, appear to be critical for silencing (Hecht et al., 1995). Deleting the highly positively charged N-terminal tail of histone H4 (Δ4-19, Δ4-23, or Δ4-28) or introducing single point mutations (K16G, K16Q, R17G, H18G, or R19G) results in derepression of the silent mating-type loci, producing a non-mating phenotype (Kayne et al., 1988) as well as loss of telomere position effect. A screen for extragenic suppressors of this phenotype produced three sir3p mutants, suggesting a direct interaction between histone H4 and Sir3p (Johnson et al., 1990). Deletion of residues 4-20 from histone H4 results in partial loss of silent mating-type silencing

and complete loss of telomeric silencing *in vivo* (Hecht et al., 1995; Thompson et al., 1994).

Interactions between Rap1p and the Sir proteins have been proposed based on the results of the two-hybrid assay. These experiments have suggested Rap1p/Sir3p, Rap1p/Sir4p, Sir3p/Sir3p, Sir3p/Sir4p, and Sir4p/Sir4p interactions (Buck and Shore, 1995; Hardy et al., 1992; Liu and Lustig, 1996). Though no Rap1p/Rap1p interactions were detected, it is important to note that in all cases where Rap1p was the "bait", at most the C-terminal 200 residues were included (Moretti et al., 1994).

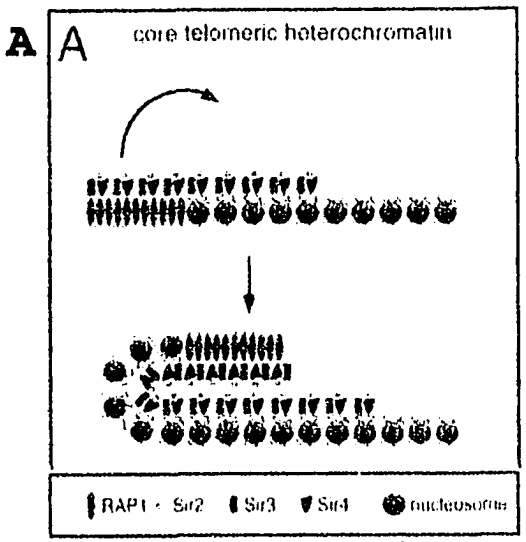
A number of direct interactions of the protein components of telomeres and silent mating-type loci have also been detected biochemically. Sir2p and Sir3p are specifically retained by an affinity column composed of the C-terminal half of Sir4p (Moazed and Johnson, 1996). When Sir4p was expressed as a GST fusion in yeast, a Sir2p/Sir4p complex was coimmmunoprecipitated from yeast extracts. However, when only a fragment containing the C-terminal 244 residues of Sir4p was expressed, a Sir2p/Sir3p/Sir4p complex was coimmmunoprecipitated,

suggesting that the Sir4p N-terminus inhibits the Sir3p/Sir4p interaction. Relief of this inhibition may be a regulated step in the assembly of silent chromatin. Results from affinity chromatography using proteins purified from *E.coli* suggest the following direct interactions: Sir2p/Sir2p, Sir2p/Sir3p, Sir2p/Sir4p, Sir3p/Sir4p, Sir4p/Sir4p, and Sir3p/Rap1p (Moazed et al., 1997). Western blotting of epitope-tagged Sir3p immunoprecipitated from whole cell extracts shows that Rap1p, Sir4p, and histones H2A, H2B, H3 and H4 all coimmunoprecipitate with Sir3p. After DNaseI digestion of this complex, Sir4p remains associated with Sir3p and the histones remain partially associated, suggesting direct interaction. In contrast, the Rap1p/Sir3p association is completely disrupted, implying an important role for DNA in maintaining the integrity of this complex (Hecht et al., 1996).

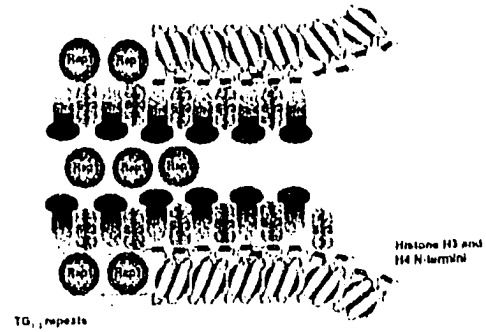
This work has lead to a number of models for the structure of the double-stranded DNA regions of yeast telomeres, four of which are reproduced in Figure 1-1. All of these models picture Rap1p primarily as a recruiter of other protein factors. The Gotta and Gasser

model (Figure 1-1 B) suggests a role for Rap1p in mediating telomere-telomere interaction, in a Sir3p-Sir4p-dependent manner. None of these models account for the evidence that Rap1p is the primary structural protein at yeast telomeres (Wright and Zakian, 1995).

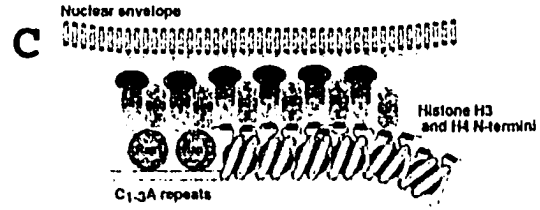
The work described here has focused on Rap1p which interacts with DNA and directs assembly of the other factors onto specific locations on the DNA. Rap1p is an 827 amino acid protein containing three functional regions defined by deletion and mutational analysis (Hardy et al., 1992; Henry et al., 1990). These are the N-terminal region, the DNA binding region and the C-terminal region. The Rap1 DNA binding region includes 235 amino acid segment spanning residues 361-596 (Henry et al., 1990). The structure of the DNA binding domain (residues 353 to 598) complexed with DNA has been solved by x-ray crystallography at 2.25Å resolution (Konig et al., 1996). The DNA binding domain contains two repeats of a protein fold that resembles the homeodomain and the proto-oncogene Myb. These two Myb-like folds bind in a tandem orientation to a tandemly repeated DNA sequence within the Rap1 binding site (Konig et al., 1996).



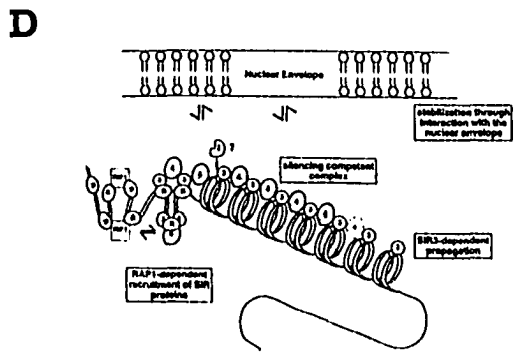
Strahl-Bolsinger
et al, Genes Dev,
11, 83-93 (1997)



Gotta and Gasser
Experientia 52, 1136-1145
(1996)



Hecht *et al*
Cell 80, 583-592
(1995)



Cockell et al. Journal of
Cell Biology 129, 909-924
(1995)

Figure 1-1 shown are four models for the structure of the double-stranded region of yeast telomeres. These represent the prevailing view that Rap1p is a monomer. No Rap1p-Rap1p interactions are proposed. Rap1p is seen primarily as a recruiter of Sir3p and Sir4p. In the model shown in (B), Gotta and Gasser propose a role for Rap1p in mediating telomere-telomere interactions.

The Rap1p C-terminal region has been extensively characterized using mutagenesis and deletion analysis. Residues 630-695 are involved in the Rap1p transactivation function, whereas residues 662-827 play a role in *HML*, *HMR* and telomeric silencing and in telomere length regulation (Buck and Shore, 1995; Gilson and Gasser, 1995; Goncalves et al., 1996; Hardy et al., 1992; Hardy et al., 1992; Kyrion et al., 1992; Liu et al., 1994; Moretti et al., 1994; Sussel and Shore, 1991; Wotton and Shore, 1997). Two classes of Rap1p C-terminal mutations have been generated. The first includes a temperature-sensitive mutation, *rap1-5* (P694L) which, at the semipermissive temperature, fails to activate *MAT*α, and results in decreased telomere tract length and derepression at *HMR***a** in a strain with a partially disabled silencer (Kurtz and Shore, 1991). Another *rap1*^{ts} allele, *grc4* also shows decreased telomere tract length (Conrad et al., 1990).

All other mutations and deletions within the C-terminal domain of Rap1p exhibit derepression of the silent mating-type loci and/or telomere-proximal genes as

well as increased telomere tract length. Sussel and Shore generated *rap1* alleles, designated *rap1^s*, specifically defective in silencing, but fully functional as activators of *MAT*α expression. This demonstrated that Rap1p activation and silencing functions are separable (Sussel and Shore, 1991). Lustig and colleagues developed a series of alleles with mutations in the C-terminal-most 28 residues of Rap1p whose phenotypes include loss of telomeric silencing and increased telomere tract length. Overexpression of Sir3p rescued telomeric silencing but not telomeric tract length. Overexpression of the C-terminal region of Rap1p (*rap1Δ19-497*) results in telomere lengthening, probably due to titration of factors important for telomere length regulation (Conrad et al., 1990). Together these results strongly suggest that the C-terminal domain of Rap1p is important for interactions with other protein factors. This conclusion is also supported by the genetic and biochemical studies discussed earlier which show that the C-terminal domain of Rap1p interacts directly with Sir3p and Sir4p.

The Rap1p N-terminal region of ~350 residues, is

highly negatively charged and has been described as dispensable for function. Yeast strains bearing *rap1Δ43-279* are viable and have no reported phenotypic consequences (Gilson and Gasser, 1995; Moretti et al., 1994). *In vitro*, the N-terminal domain appears to be required for Rap1p DNA bending activity (Muller et al., 1994; Vignais et al., 1989).

Several lines of evidence lead to the expectation that phosphorylation will be found to play a role in some aspect of Rap1p function. A search for suppressors of the toxicity of Rap1p overexpression yielded Sko1p/Acr1p. This protein had been previously identified as a repressor of toxicity caused by overexpression of a cAMP dependent protein kinase (Freeman et al., 1995). It is interesting that growth-regulated induction of ribosomal protein genes depends strongly on the ability to regulate protein kinase A (a cAMP-dependent kinase) and that this PKA-dependent induction requires Rap1p binding sites (Klein and Struhl, 1994). Putting this together with the result that Sko1p/Acr1p suppresses toxicity from overexpression of both Rap1p and a cAMP-dependent protein kinase suggests that phosphorylation of Rap1p or of an

abundant protein that interacts with Rap1p may be required for activation of ribosomal protein genes. It has been reported that binding of Rap1p from whole cell extracts to a labeled PGK UAS (glycolytic enzyme) probe is carbon-source dependent and that the DNA binding affinity is increased by treatment with phosphatase (Tsang et al., 1990).

This research on Rap1p has focused on biochemical studies that lead to a more precise understanding of its solution properties and DNA binding activities *in vitro* in preparation for structural studies. The work described here includes 1) overexpression and characterization of full length Rap1p and its structural domains, 2) crystallization of Rap1 domains alone and complexed with DNA, and 3) the discovery of a function for the N-terminus of Rap1p which involves oligomerization of Rap1p.

Chapter II

Large Scale Production of Rap1p in *E. coli* and Characterization of Its DNA Binding Properties and Domain Structure

Introduction

The first requirement for biophysical and biochemical studies of Rap1p was a source of large quantities of highly purified protein. Previous purification from yeast yielded only 10-20 μg from 100 L of yeast culture (Shore and Nasmyth, 1987). *E. coli* lysates from lysogens of λgt11 clones expressing Rap1p were used in gel shift experiments, indicating that Rap1p expressed in *E. coli* behaved similarly to that expressed in yeast (Shore and Nasmyth, 1987), but no overexpression and purification from *E. coli* had been reported when we began this work. We constructed a Rap1p expression plasmid, overexpressed Rap1p in *E. coli* and developed an efficient purification scheme that yields greater than 10

mg of highly purified Rap1p from 3L of *E. coli* culture. Using this material, we went on to evaluate the binding properties and domain structure of Rap1p.

Rap1p was identified as a protein that binds to DNA sequences found in the yeast mating-type silencers (Buchman et al., 1988; Shore et al., 1987), the upstream activating sequence required for transcription of the *MAT*α genes (Buchman et al., 1988), the upstream promoter elements of ribosomal protein genes (Buchman et al., 1988; Huet et al., 1985), and telomeric $(C_{1-3}A)_n$ repeats (Berman et al., 1986; Buchman et al., 1988). A compilation of the naturally occurring Rap1p binding sites led to the consensus sequence listed in Figure 2-1 (Buchman et al., 1988). It was reported that Rap1p purified from yeast binds with a dissociation constant of 13 pM when a composite RPG sequence was used as the probe (Vignais et al., 1990). Buchman *et al* reported that Rap1p may have a slightly higher affinity (less than 2-fold) for the telomeric repeat (Buchman et al., 1988).

A slightly extended and more constrained Rap1p binding site was identified using a selection technique referred to as selected and amplified binding (SAAB).

The starting material was a 57 base pair oligonucleotide containing a 13 base pair region of total degeneracy. DNA was selected for its ability to bind Rap1p produced by *in vitro* translation. After two rounds of selection, 47 sequences were determined and the "selected" Rap1p consensus binding site was defined (see Figure 2-1) (Graham and Chambers, 1994).

From yeast sites A/G A/C A C C C A N N C A C/T C/T
From selected sites A/G T A/G C A C C C A N N C C/A C C

Figure 2-1 Rap1p Consensus Sequences

Deletion analysis demonstrated that essentially all of the sequence-specific DNA binding properties of Rap1p reside in the central region of Rap1p (residues 361-596) (Henry et al., 1990). A search for amino acid sequence homologies did not reveal any of the known DNA binding motifs, but subsequent structure determination revealed that the Rap1p DNA binding domain contains two repeats of a protein fold that resembles the homeodomain and the proto-oncogene Myb (Konig et al., 1996).

The C-terminal region of the protein contains overlapping transcriptional activation (residues 630-695)

and silencing (residues 678-827) regions as demonstrated by deletion and mutational analysis (Hardy et al., 1992). Mutations in the C-terminal region also affect telomere length regulation (Kyrion et al., 1992; Lustig et al., 1990).

The N-terminal region was found to be dispensable for viability (Shore, 1994) and its deletion has been reported to have no effect on silencing at the silent mating-type loci or at telomeres (Moretti et al., 1994). DNA bending by Rap1p *in vitro*, as evaluated by proteolysis experiments (Vignais et al., 1989), circular permutation, and electron microscopy is reported to require the N-terminal domain (Muller et al., 1994).

Using limited proteolysis, we defined three stable structural domains within Rap1p: an N-terminal domain, a central DNA binding domain, and a C-terminal domain that contains the residues important for activation, silencing and telomere length regulation. These domains were subcloned, overexpressed, and purified for utilization in our crystallization efforts and for our biochemical studies.

Materials and Methods

Expression and Purification of Rap1p

*Eco*R I and *Sal* I restriction sites were introduced immediately before the ATG start codon of *RAP1* and immediately following the TAA stop codon respectively by PCR using the following oligonucleotides:

RAPNT CCT TCC TTC CTT GAA TTC ATG TCT AGT CCA GAT T

RAPCT2 CCT CTC CCT TTC GTC GAC TTA TAA CAG GTC CTT CTC AAA AAA.

The PCR fragment was digested with *Eco*R I and *Sal* I and subcloned into pKKT7E to create plasmid pKKT7E-RAP1. The *Pst* I-*PpuM* I fragment of the *RAP1* gene was removed and replaced with the same fragment from pLP11 to eliminate mutations that may have been introduced by PCR. The plasmid was introduced into *E. coli* BL21(DE3) pLysS cells by electroporation. Cells were grown in prewarmed 2XYT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl) containing 30 mM Kpi (pH 7.8), 5 mM glucose, and 500 µg/ml ampicillin. The cells were grown at 37°C to an OD_{600} of 0.5 and then cooled to room temperature and grown to an OD_{600} of 1.0. Rap1p production was induced by

addition of IPTG to a final concentration of 0.5 mM and the cells were incubated overnight at room temperature. Cells were harvested by centrifugation and the pellets were frozen at -15•C.

The cells (~30 grams, wet weight) were resuspended in 300 ml of buffer I (50 mM Tris (pH 7.5), 50 mM NaCl, 0.02% NaN_3, 1 mM EDTA, 2 mM DTT) and lysed by sonication. Cell debris was removed by centrifugation for 30 minutes at 24,000 x g and the supernatant was collected. Solid ammonium sulfate was added to a final concentration of 40% of saturation and the solution was stirred on ice for 30 minutes. The precipitate, which contained the Rap1p, was collected by centrifugation and the pellet was resuspended in 30 ml buffer II (25 mM Hepes (pH 7.5), 100 mM NaCl, 0.02% NaN_3, 0.25 mM EDTA, 2 mM DTT) and dialyzed overnight against 2 L of the same buffer. Precipitate was removed by centrifugation for 30 minutes at 39,000 x g and the supernatant was filtered through 0.45 µm and 0.22 µm filters. This filtered solution was loaded onto a HiLoad 16/10 S-Sepharose column (Pharmacia) using a Pharmacia FPLC system. Rap1p was eluted using a gradient of 0-30% buffer III (25 mM Hepes (pH 7.5), 1.0 M NaCl,

0.02% NaN_3, 0.25 mM EDTA, 2 mM DTT) over 60 minutes. Fractions containing Rap1p were pooled and dialyzed overnight against buffer IV (25 mM Tris (pH 7.5), 100 mM NaCl, 0.02% NaN_3, 0.1 mM EDTA, 2 mM DTT). This solution was then filtered, loaded onto a HiLoad 16/10 Q-Sepharose column (Pharmacia) and eluted with a gradient of 0-35% buffer V (25 mM Tris (pH 7.5), 1.0 M NaCl, 0.02% NaN_3, 0.1 mM EDTA, 2 mM DTT) over 60 minutes. Fractions containing Rap1p were pooled and concentrated to less than 500 µl using Centriprep 30 and Centricon 30 spin concentrators (Amicon). The protein was then run over a Superdex 75 gel filtration column (Pharmacia) in buffer VI (5 mM Hepes (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 0.1 mM EDTA, 2 mM DTT) and stored at 4°C. Typical yields were 10 mg of greater than 98% pure Rap1p from a 3 L cell growth (~30 grams of cells, wet weight).

Limited Proteolysis

Rap1p (1 mg/ml) was incubated in buffer VI (5 mM Hepes (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 2 mM DTT) at 4°C with either α-chymotrypsin or trypsin at a protein:protease ratio of 800:1 (w:w). Aliquots were

removed from the reaction at various times and protease activity was stopped by addition of SDS gel loading buffer (10% glycerol, 5% β-mercaptoethanol, 3.05% SDS, 62.5 mM Tris (6.8), followed by boiling for 5 minutes. These samples were analyzed by SDS polyacrylamide gel electrophoresis.

For N-terminal sequencing and western blotting, the proteins were transferred to Immobilon as follows. The gels were equilibrated in transfer buffer (48 mM Tris; 39 mM glycine; 0.375% SDS; 20% methanol). Immobilon was pre-wet in 100% methanol, soaked briefly in water, and then equilibrated in transfer buffer. Transfer of proteins was done using a BIO-RAD Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell according to manufacturer's instructions. The western blots utilized an antibody to the C-terminus of Rap1p (a generous gift of Lorraine Pillus). The antibody was raised in rabbit to the following peptide (which corresponds to the last 14 residues of Rap1p) conjugated to Keyhole-limpet hemocyanin via *m*-Maleimidobenzoyl-N-hydroxysuccinimide ester: CRMEMRKRFFEKDLL. For sequencing, the Immobilon was treated with coomassie blue (30% methanol; 0.25%

Coomassie Blue R) and bands were sliced out with a razor blade. N-terminal sequencing was performed by Macromolecular Resources, Fort Collins, CO.

Expression and Purification of Rap1p Structural Domains

Rap1 DNA Binding Domain

The following oligonucleotides were used as PCR primers to amplify DNA fragments encoding the DNA binding domain of Rap1p.

RAPDNANTI CCT TCC TTC CTT GAA TTC ATG ATT TCG AGA GGC GCT

RAPDNANTII CCT TCC TTC CTT GAA TTC ATG AAT GCA TCT TTT CAA GCA CAA

RAPDNACTERM CCT CTC CCT TTC GTC GAC TTA AGA ACT ATA ATT TCT TGC CCT

The first primer (RAPDNANTI) introduces an *Eco*R I restriction site immediately 5' to the codon for M349 and the second primer (RAPDNANTII) introduces an *Eco*R I restriction site and an ATG codon immediately 5' to the codon for N340. The RAPDNACTERM primer introduces a TAA stop codon and a *Sal* I restriction site immediately 3' to the codon for S604.

The fragments were amplified using Taq polymerase, digested with *Eco*R I and *Sal* I, and ligated into pKKT7E to create plasmids pKKT7E-Rap1DBD-A and pKKT7E-Rap1DBD-B (RAPDNANTI and RAPDNANTII, respectively).

Rap1 DNA Binding Domain plus C-terminal Domain

The following oligonucleotides were used as PCR primers to amplify DNA fragments encoding the DNA binding domain plus C-terminal domain of Rap1p.

RAPDNANTI CCT TCC TTC CTT GAA TTC ATG ATT TCG AGA GGC GCT

RAPDNANTII CCT TCC TTC CTT GAA TTC ATG AAT GCA TCT TTT CAA GCA CAA

RAPCT2 CCT CTC CCT TTC GTC GAC TTA TAA CAG GTC CTT CTC AAA AAA.

The first primer (RAPDNANTI) introduces an *Eco*R I restriction site immediately 5' to the codon for M349 and the second primer (RAPDNANTII) introduces an *Eco*R I restriction site and an ATG codon immediately 5' to the codon for N340. The RAPCT2 introduces a SalI immediately 3' to the TAA stop codon. The fragments were amplified using Taq polymerase, digested with *Eco*R I and *Sal* I, and ligated into pKKT7E to create plasmids pKKT7E-Rap1DBDTEL-

C and pKKT7E-Rap1DBDTEL-D (RAPDNANTI and RAPDNANTII, respectively).

Purification of Rap1 DBD and Rap1 DBDTEL proteins

Rap1DBD and Rap1DBDTEL are purified using the same basic protocol as for the full length protein. Briefly, the protein is purified by ammonium sulfate precipitation, cation exchange chromatography, anion exchange chromatography, and, finally, gel filtration chromatography. We found that Rap1DBD and Rap1DBDTEL are noticeably less soluble at high protein concentrations and low salt than Rap1p. Therefore, all solutions contain at least 0.1 M NaCl and 0.2 NaCl is used in the gel filtration step. Purified protein is stored in 0.2 M NaCl. Typical yields are 3-5 mg of Rap1DBD and 10 mg of Rap1DBDTEL from 3 L of cells.

RAP1 N-terminal Domain

An *Eco*R I restriction site was introduced immediately before the *RAP1* ATG start codon and a TAA stop codon and a SalI restriction site were introduced immediately after the codon for A345, by PCR using following oligonucleotides:

RAPNT CCT TCC TTC CTT GAA TTC ATG TCT AGT CCA GAT T

RPNTCT CCT CTC CCT TTC GTC GAC TTA TGC TTG AAA AGA TGC ATT CC

The PCR fragment was digested with *Eco*R I and *Sal* I and subcloned into pKKT7E to create plasmid pKKT7E-RAP1NTD.

Rap1 N-terminal Domain plus DNA Binding Domain

An *Eco*R I restriction site was introduced immediately before the *RAP1* ATG start codon and a TAA stop codon and a *Sal* I restriction site were introduced immediately after the codon for S604 by PCR using following oligonucleotides:

RAPNT CCT TCC TTC CTT GAA TTC ATG TCT AGT CCA GAT T

RAPDNACTERM CCT CTC CCT TTC GTC GAC TTA AGA ACT ATA ATT TCT TGC CCT

The PCR fragment was digested with *Eco*R I and *Sal* I and subcloned into pKKT7E to create plasmid pKKT7E-RAP1 NTD+DBD.

Purification of Rap1 NTD and Rap1 N+DBD

Purification of Rap1 NTD and Rap1 NTD+DBD differs slightly from that of full length Rap1p. These proteins are found in the precipitate after addition of 30%-60% saturated ammonium sulfate rather than in less than 40%

SAS as for Rap1p, Rap1 DBD and Rap1 DBDTEL. The remainder of the purification of Rap1 NTD+DBD is identical to that described above for Rap1p, Rap1 DBD and Rap1 DBDTEL. However, Rap1 NTD is not retained on a cation exchange column and is found in the flow-through fractions. Rap1 NTD is eluted from the anion exchange column similarly to full length Rap1p.

Binding Studies

DNA binding assays for Rap1p and structural domains were performed by native gel electrophoresis as follows. Protein was diluted into binding buffer (50 mM Hepes (pH 7.5), 80 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 5.0% glycerol, 100 •g/ mL BSA) then added to an equal volume of the same buffer containing 2.0 pM ^{32}P-labeled DNA and incubated at room temperature for 24 hours. Bound DNA was separated from unbound by polyacrylamide gel electrophoresis, (7%, 99:1 acrylamide:bisacrylamide for Rap1p and Rap1 NTD+DBD, 10%, 55:1 acrylamide:bisacrylamide for Rap1 DBD and Rap1 DBDTEL; 1x TBE). Gels were run at 10 mA for ~4 hours, dried, visualized using a PhosphorImager (Molecular Dynamics), and quantified using ImageQuant (Molecular

Dynamics). The data were plotted and analyzed using KaleidaGraph (Synergy Software). The data were fit to equation 1.

Equation 1 $Y=((M0^{M1})/(M2+(M0^{M1}))$

M0=free protein concentration

M1=Hill coefficient

$M2=K_d$

Experiments were performed at least three times using independently prepared protein dilutions. Standard errors of fraction bound were calculated and used as estimated errors.

The data reported in this chapter were collected from experiments using the following DNA (the Rap1p binding site is in bold):

RB33mer

AAT TCA CAC CC**A CAC CCA CAC ACC** ACA CCA CAG
GT GTG GG**T GTG GGT GTG TGG** TGT GGT GTC CTA G

RESULTS

Rap1p Purification

Coomassie blue stained SDS polyacrylamide gels that show the expression and purification of full length Rap1p from *E. coli* are shown in Figure 2-2. A yield of 10 mg of protein can be obtained from 3 L of cell culture using our expression/purification system. This protein is very soluble, even in low salt solutions (>10 mg/ml in 50 mM NaCl), and is stable for >3 weeks when stored at 4°C in sizing column buffer.

Identification of structural domains

Purified Rap1p was subjected to proteolysis with trypsin, which cleaves C-terminal to lysine and arginine residues, or α-chymotrypsin, which cleaves C-terminal to large hydrophobic residues. The reactions were carried out at 4°C, using a protein:protease ratio of 800:1. Higher temperatures and higher protease concentrations yielded the same fragments, but these fragments were more rapidly degraded. Time courses for the trypsin and α-chymotrypsin trials are shown in Figure 2-3. The identities of the fragments were determined by N-terminal

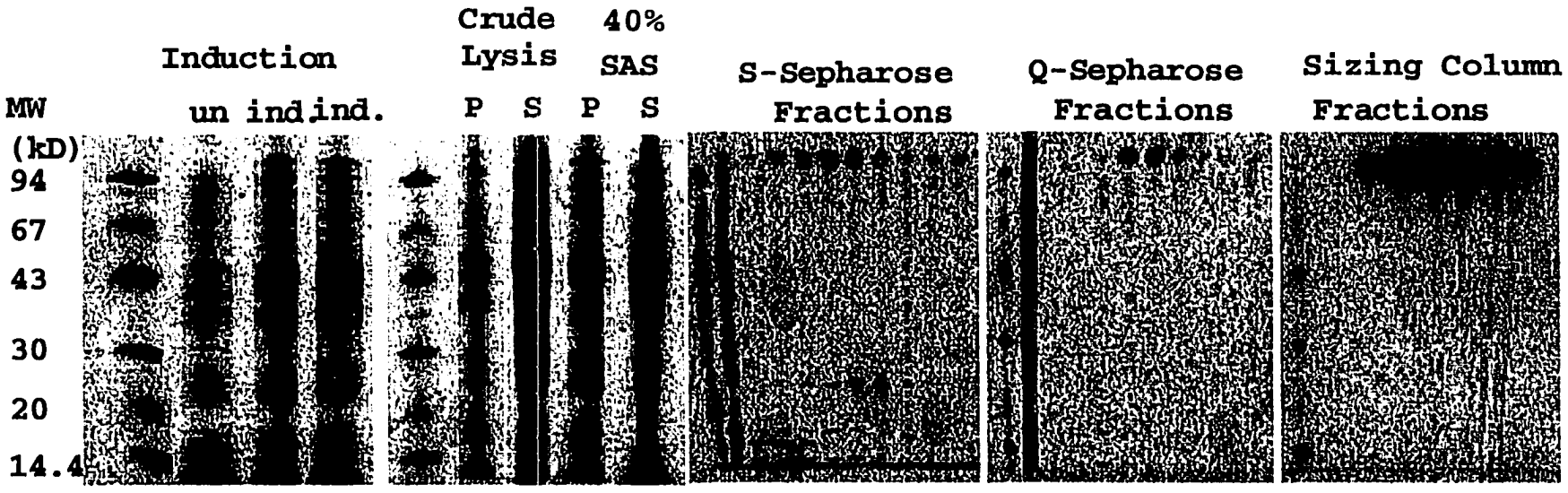


Figure 2-2. RAP1 Purification Shown are SDS polyacrylamide gels of Rap1p induction and purification. Rap1p is a 92.5 kD protein that runs at ~116 kD, indicated by the arrow. On the induction gel, Rap1p is visible in the two induced lanes (ind). Rap1p is found in the supernatant after lysis and is precipitated in 40% saturated ammonium sulfate (SAS). Purification on S-sepharose, Q-sepharose and Sizing columns yields >98% pure Rap1p.

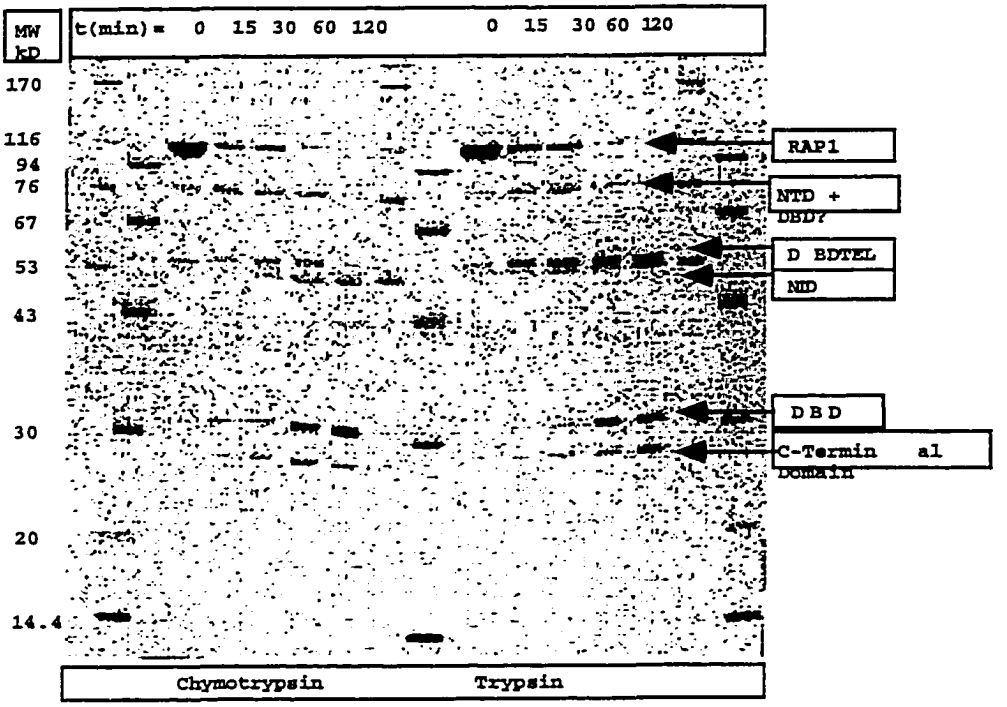


Figure 2-3 Limited Proteolysis of RAP1 Reveals 3 Stable Structural Domains Shown here are the results of time courses of limited proteolysis of Rap1p with chymotrypsin and trypsin. Both proteases yield 3 stable structural domains: the N-terminal domain, the DNA binding domain, and the C-terminal domain.

sequencing and western blotting with an antibody to the C-terminus of Rap1p. Results of N-terminal sequencing and recognition by an antibody to the C-terminus of the protein are shown in Figures 2-4 and 2-5.

Trypsin was found to cleave immediately C-terminal to R352 and R606, while α-chymotrypsin cleaves C-terminal to F343 and Y602. These results are summarized in Table 2.1 and shown schematically in Figure 2-6.

Four structural domains of Rap1p were subcloned, overexpressed and purified. These are Rap1 N-terminal domain, residues M1-A345, (Rap1 NTD); Rap DNA binding domain, residues N340 or M349-S604, (Rap1 DBD); Rap1 N-terminal domain plus DNA binding domain, residues M1-S604 (Rap1 NTD+DBD); and Rap1 DBD binding domain plus C-terminal domain, residues N340 or M349-L827, (Rap1 DBDTEL). These are shown schematically in Figure 2-7.

DNA binding properties

Examples of gel shift experiments for Rap1p, Rap1 DBD, Rap1 DBDTEL, Rap1 N-D, and Rap1 NTD are shown in Figure 2-8. Native polyacrylamide gel electrophoresis was used to separate bound from unbound DNA. The DNA concentration was 1 pM, which is 20 times lower than the

```
Fragment submitted:Rap1p
Sequence obtained                  P D P F E
Rap1p sequence            1 M S S P D D F E T A 10
```

Protease: Trypsin

```
Fragment submitted:50 KD Lower
Sequence obtained                  P D - F E N A P A E
Rap1p sequence            1 M S S P D D F E T A P A E Y 14


Fragment submitted: 30 kD Upper
Sequence obtained                    H A R P F A
                                       P L   K R
Rap1p sequence            350 I S R G A L P S H N K 361

Fragment submitted: 30 kD Lower
Sequence obtained                        N V Q P P A N
Rap1p sequence      600 R N Y S S Q R N V Q P T A N A 614
```

Protease: α-chymotrypsin

```
Fragment submitted: 50 kD Lower
Sequence obtained                  P N P F E
                                   K   D
Rap1p sequence            1 M S S P D D F E T A P A E Y 14

Fragment submitted: 30 kD Upper
Sequence obtained                      Q A Q R T
Rap1p sequence            340 N A S F Q A Q R S M I S R
352


Fragment submitted: 30 kD Lower
Sequence obtained                  - - Q R N V Q
Rap1p sequence        600 R N Y S S Q R N V Q P T A N A
614
```

Figure 2-4 N-terminal sequences obtained for full length Rap1p and proteolytic fragments from limited digestion with α-chymotrypsin and trypsin. Each fragment is identified by its approximate size on an SDS polyacrylamide gel (Figure 2-3), the sequence obtained and the alignment with the Rap1p amino acid sequence.

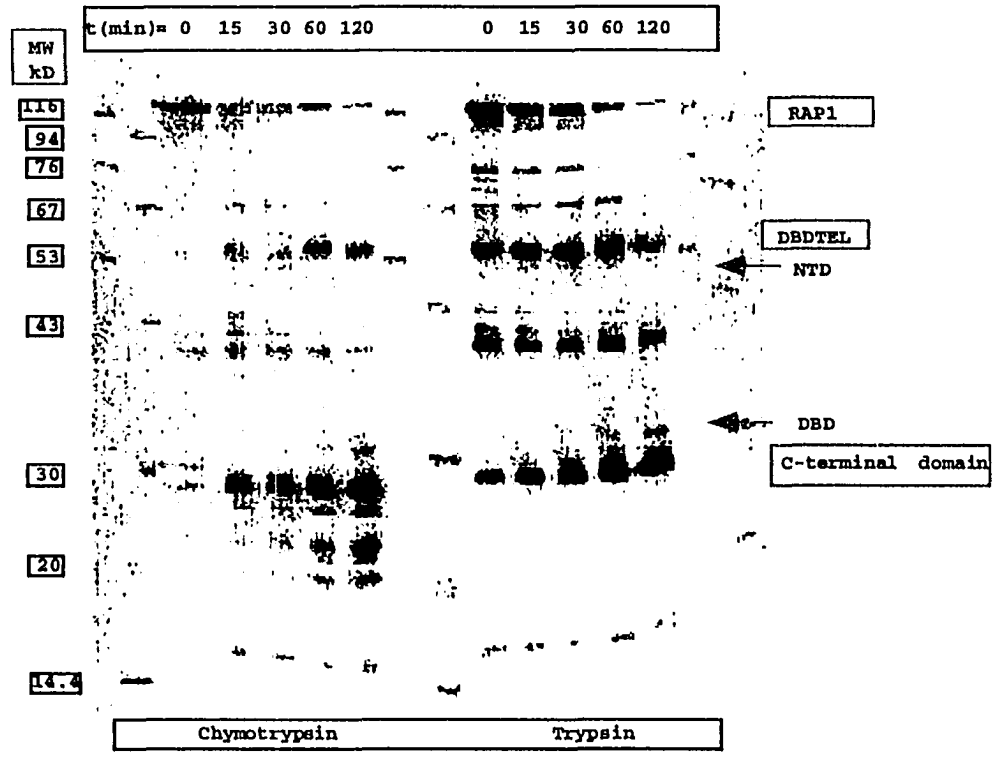


Figure 2-5 Western Blot of Limited Proteolysis Gel Using Antibody to Rap1p C-terminus Shown here is a western blot of the gel pictured in Figure 2-3. Proteins have been transferred to Immobilon using electrophoresis and the membrane has been probed with antibody to the C-terminus of Rap1p. Structural domains that are recognized by the antibody are Rap1p, the DNA binding plus C-terminal domain (DBDTEL), and the C-terminal domain. The N-terminal domain and DNA binding domain (marked with arrows) are not recognized by the antibody.

Table 2.1 Summary of Determination of Rap1p Structural Domains Stable proteolytic fragments of Rap1p were identified by determining their N-terminal sequences and recognition by an antibody to the C-terminus of Rap1p.

Apparent MW of Stable fragments	N-terminus after cleavage with		Recognized by C-terminal antibody	Identity
	chymotrypsin	trypsin		
50 kD Upper	F343	R352	Yes	Rap1DBDTEL Residues ~350-827
50 kD Lower	M1	M1	No	Rap1NTD Residues 1-~350
30 kD Upper	ND	ND	No	Rap1DBD Residues~350-~600
30 kD Lower	Y602	R606	Yes	Rap1TEL Residues ~600-827

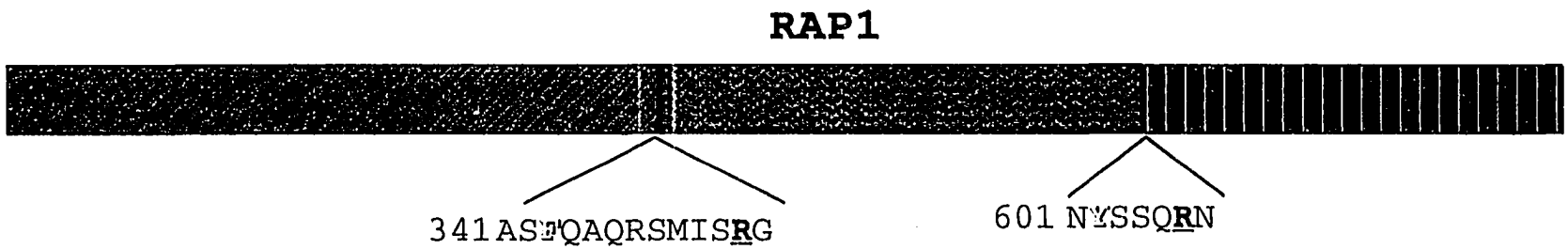


Figure 2-6 Schematic representation of Rap1p showing points of cleavage by a-chymotrypsin (F343 and Y602) and trypsin (R352 and R606).

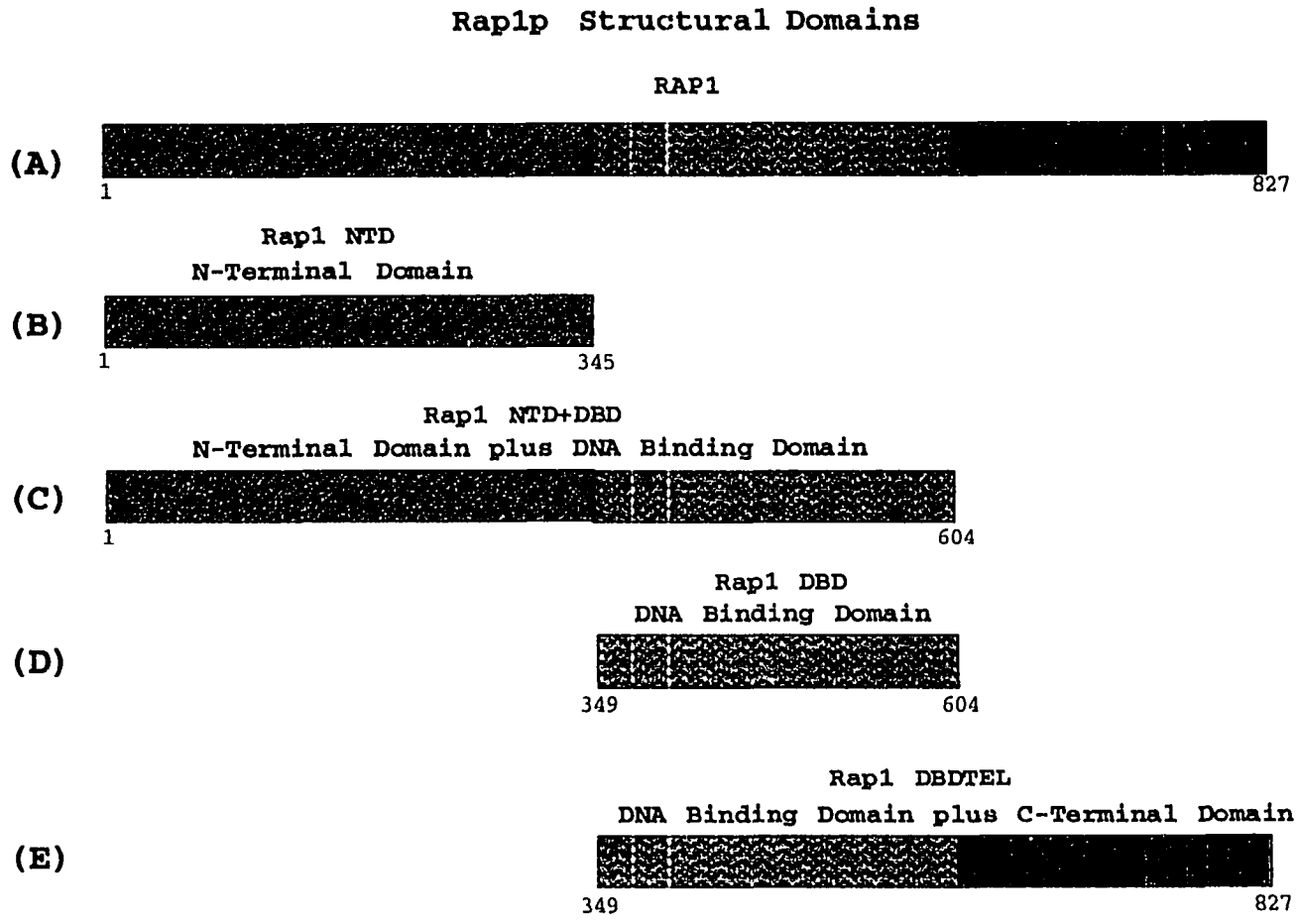


Figure 2-7 Shown are schematic representations of (A)full length Rap1p, (B) Rap1 N-terminal Domain, (C) Rap1 N-terminal Domain plus DNA binding Domain, (D) Rap1 DNA Binding Domain, and (E) Rap1 DNA Binding Domain plus C-terminal Domain. These constructs were subcloned, overexpressed in *E. coli*, and purified for use in crystallization trials and biochemical studies.

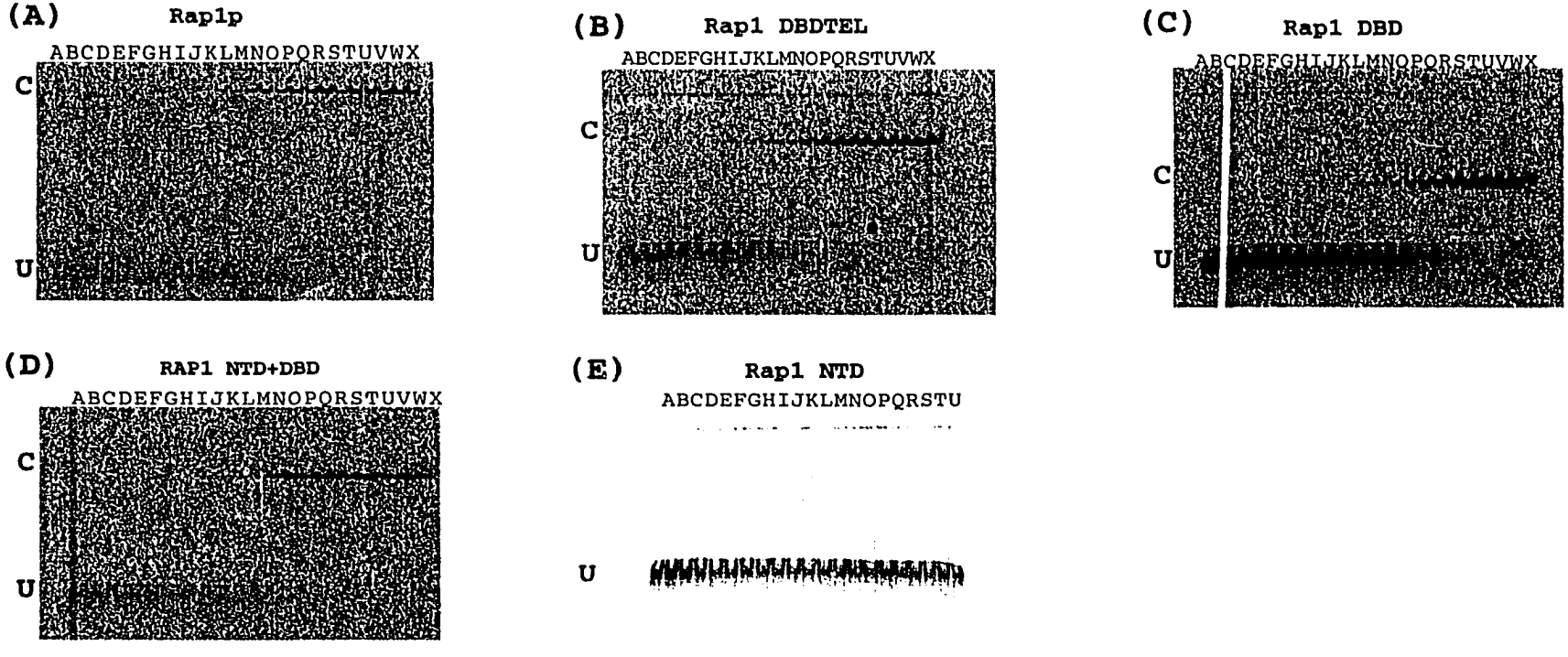


Figure 2.8 Native polyacrylamide gels showing binding of (A) DBD, (B) N + DBD, (C) DBDTEL, (D) Rap1p and (E) NTD to DNA. A,B,C and D bind with similar affinities. NTD does not bind DNA. Unbound DNA (U) and DNA complexed with protein (C) are indicated.

Protein concentrations (X 10^{-10} M) for gels (A), (B), (C), and (D) are as follows.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
0	.01	.015	.02	.03	.04	.06	.09	.12	.17	.24	.35	.50	.71	1	1.4	2.1	2.9	4.2	6	8.6	12	18	25

Protein concentrations (X 10^{-9} M) for gel (E) are as follows.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
0	.009	.01	.017	.024	.035	.05	.07	.1	.14	.2	.3	.4	.6	.9	1.2	2.5	10	40	160	1000

lowest K_d. The conditions used for K_d determination are as follows. Protein was diluted into binding buffer (50 mM Hepes (pH7.5), 80 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 5.0% glycerol, 100 µg/ ml BSA) then added to an equal volume of the same buffer containing 2.0 pM ^{32}P-labeled DNA and incubated at room temperature for 24 hours. Rap1p, Rap1 DBD, Rap1 DBDTEL and Rap1 N-D all bind with similar K_ds between ~30 pM and 170 pM. Binding curves are shown in Figure 2-9. Binding constants are summarized in Table 2.2. We found that it was critical to have BSA in the protein dilution buffer to observe reproducibly tight binding. Diluting protein into solutions without BSA resulted in a higher K_d which steadily increased with the length of the dilution. This is illustrated in Figure 2-10. We evaluated binding at time points up to 48 hours. There is no change in K_d between 24 and 48 hours (Figure 2-11), indicating that the binding reaction has reached equilibrium after a 24 hour incubation.

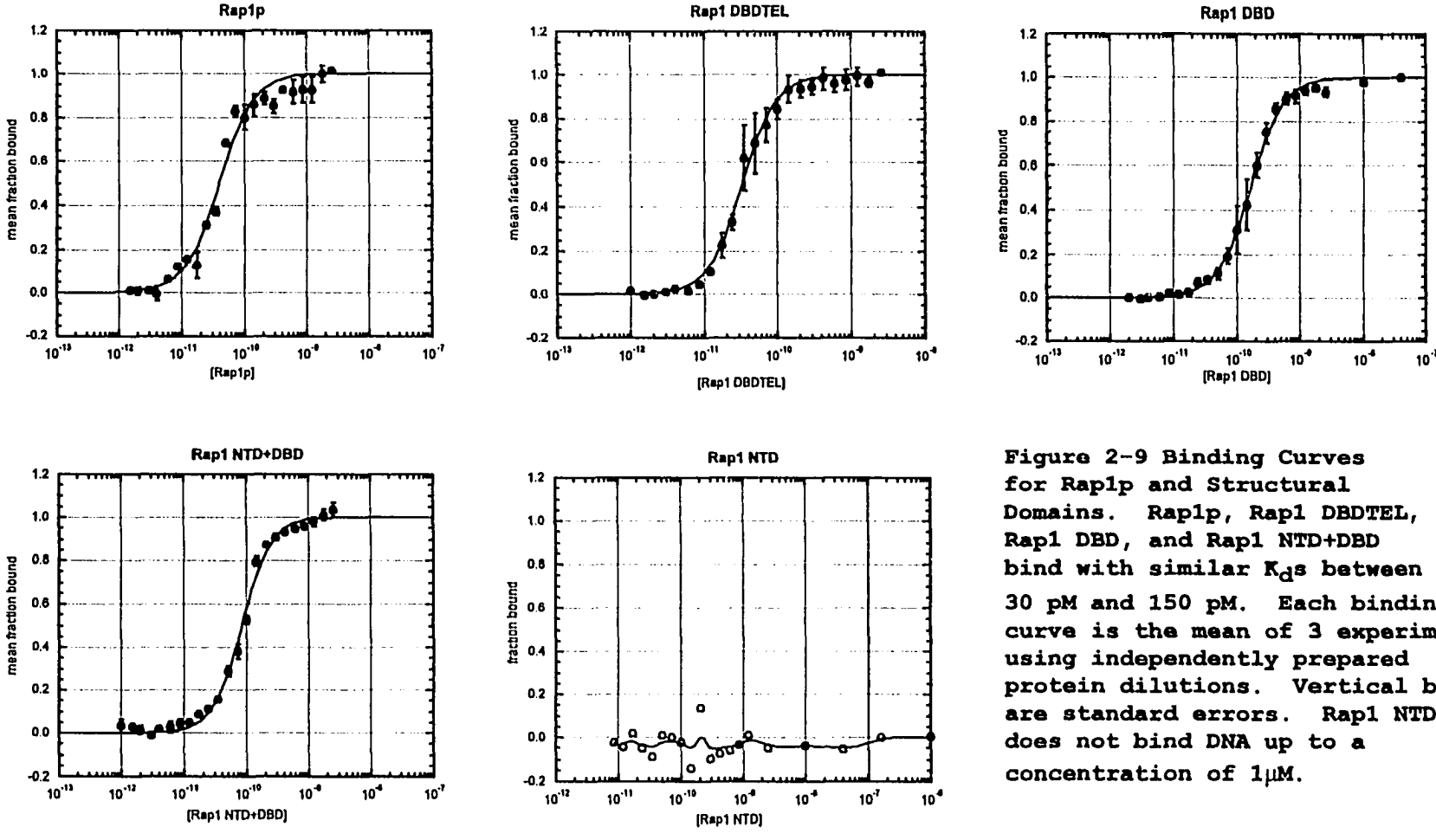


Figure 2-9 Binding Curves for Rap1p and Structural Domains. Rap1p, Rap1 DBDTEL, Rap1 DBD, and Rap1 NTD+DBD bind with similar K_ds between 30 pM and 150 pM. Each binding curve is the mean of 3 experiments using independently prepared protein dilutions. Vertical bars are standard errors. Rap1 NTD does not bind DNA up to a concentration of 1μM.

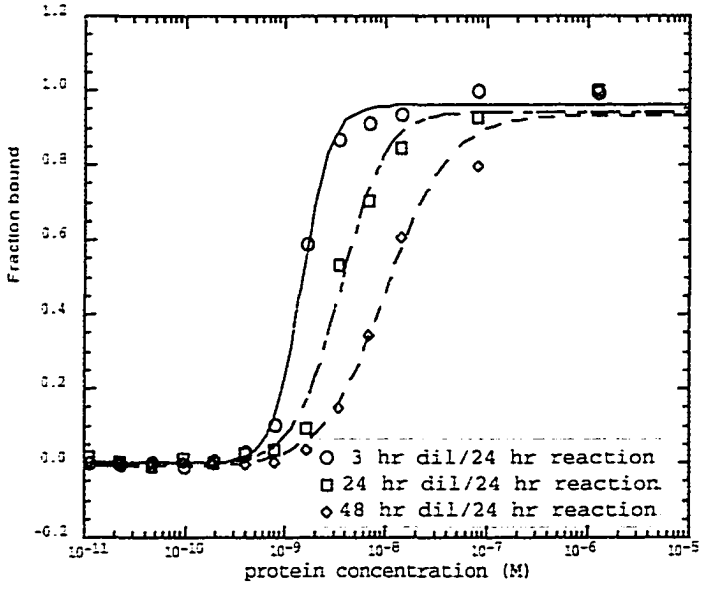


Figure 2-10 Rap1 DNA Binding domain Diluted in Sizing Column Buffer This plot shows the Kds for Rap1 DBD bound to DNA after the protein has been diluted in sizing column buffer without BSA for 3 hours, 24 hours, or 48 hours. The apparent Kd drops 1 order of magnitude over the time of the experiment from 1 nM to 10 nM.

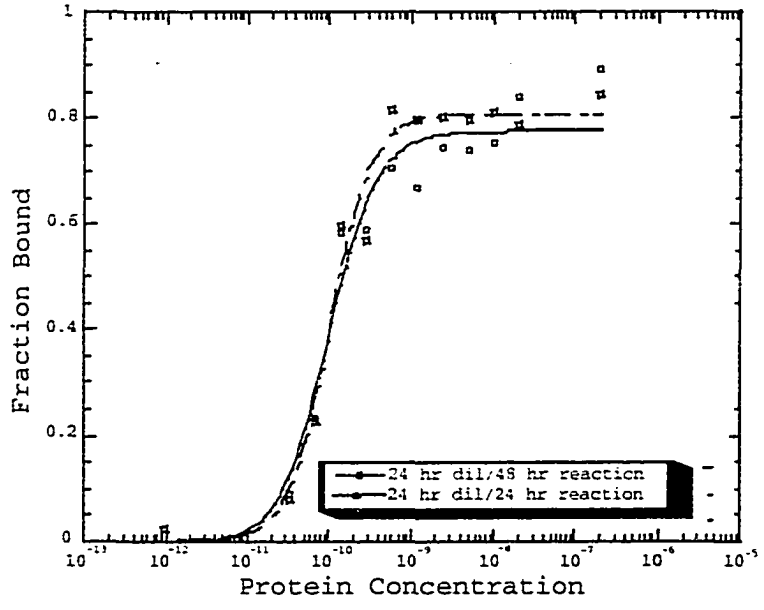


Figure 2-11 Rap1 DNA Binding Domain: 24 and 48 hour Binding Reactions Binding curves for Rap1 DBD bound to DNA for 24 and 48 hours. There is no change in Kd, indicating the reaction has reached equilibrium by 24 hours. Rap1 DBD has been diluted in binding buffer containing BSA.

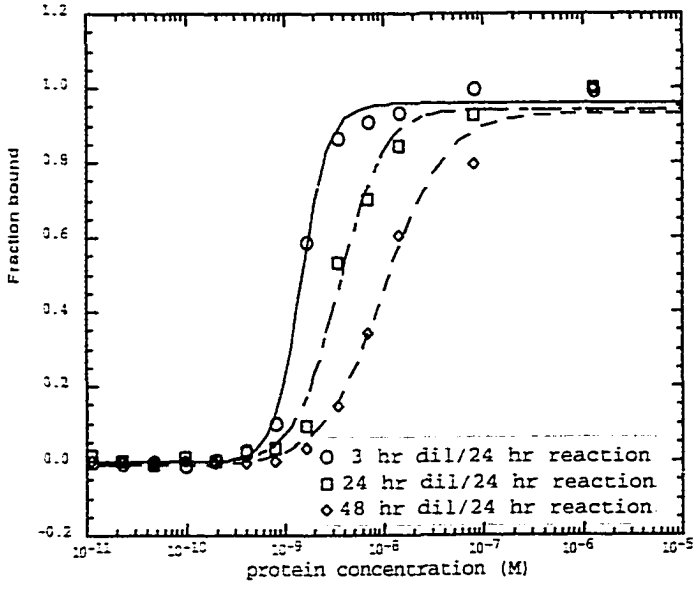


Figure 2-10 Rap1 DNA Binding domain Diluted in Sizing Column Buffer This plot shows the Kds for Rap1 DBD bound to DNA after the protein has been diluted in sizing column buffer without BSA for 3 hours, 24 hours, or 48 hours. The apparent Kd drops 1 order of magnitude over the time of the experiment from 1 nM to 10 nM.

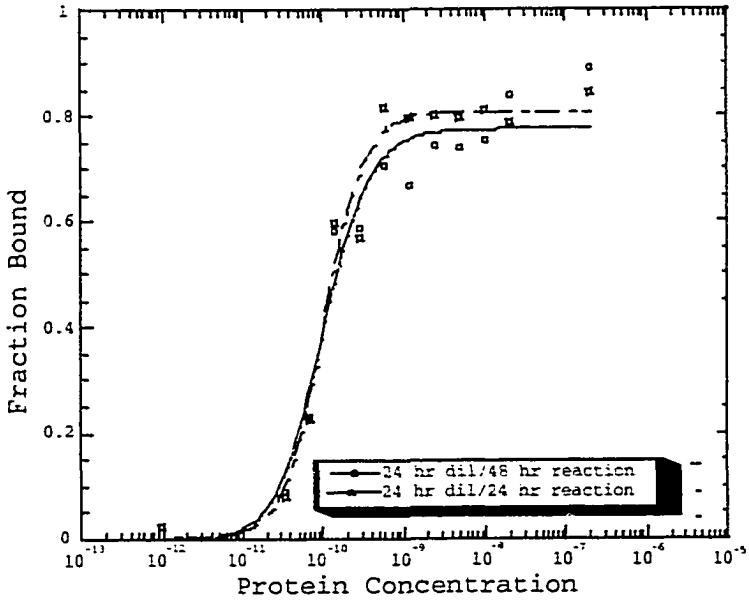


Figure 2-11 Rap1 DNA Binding Domain: 24 and 48 hour Binding Reactions Binding curves for Rap1 DBD bound to DNA for 24 and 48 hours. There is no change in Kd, indicating the reaction has reached equilibrium by 24 hours. Rap1 DBD has been diluted in binding buffer containing BSA.

Table 2.2 Dissociation Constants for Rap1p, Rap1 DBD, Rap1 DBDTEL, RAP1 NTD+DBD, and Rap1 NTD Rap1p, Rap1 DBD, Rap1 DBDTEL, RAP1 NTD+DBD have similar low picomolar k_ds, reported here with standard errors. Hill coeffients for these four proteins are very similar. Rap1 NTD does not bind DNA.

Protein	K_d (pM)	Standard Error ($x10^{-12}$)	M1 (Hill coefficient)
Rap1p	**39**	**2.7**	**1.6**
Rap1 DBD	**170**	**3.8**	**1.7**
Rap1 DBDTEL	**33**	**1.1**	**1.8**
Rap1 NTD+DBD	**84**	**2.8**	**1.8**
Rap1 NTD	**--**	**--**	**--**

Discussion

The methods we have developed for expressing and purifying large quantities of Rap1p from *E. coli* have enabled us to pursue a variety of biophysical and biochemical studies leading to a better understanding of Rap1p structure and function. To better understand the domain structure of Rap1p, we employed limited

proteolysis. Trypsin cleaved Rap1p

C-terminal to R352 and R606, while α-chymotrypsin cleaved C-terminal to F343 and Y602. The proximity of these cleavage sites by two proteases with different specificities (trypsin cuts C-terminal to lysine and arginine; α-chymotrypsin cuts C-terminal to large hydrophobic residues) is striking. Given that Rap1p contains 146 potential targets for trypsin and 89 potential targets for α-chymotrypsin, we conclude that the residues cleaved are located on highly accessible, flexible regions of the protein. Although we did not demonstrate that the fragments are separable, the behavior of the individually expressed domains indicates that these peptide segments indeed compose structural domains.

The structural domains defined by limited proteolysis correspond well to functional regions previously identified by deletion analysis (Hardy et al., 1992; Henry et al., 1990). The minimal DNA binding region is reported to encompass residues A-361 to A-596 (Henry et al., 1990). The transcriptional activation region includes residues A630 to N695, while

transcriptional silencing requires the carboxy-terminal 150 amino acids (S-678-L827) (Hardy et al., 1992). Our experiments suggests that the Rap1p DNA binding domain is a structural unit flanked by independently stable N-terminal and C-terminal domains. We went on to express and purify the following fragments of Rap1p: the DNA binding domain (339 or 349-604, the DNA binding domain plus C-terminal domain (339 or 349-827), the N-terminal domain (1-345), and the N-terminal domain plus DNA binding domain (1-604) (Figure 2-7).

The Rap1 N-terminal domain is reported to be dispensable for Rap1p function (Gilson and Gasser, 1995). As will be discussed in later chapters we have found the NTD gives rise to dimerization and tetramerization of Rap1p and we report a phenotype in yeast containing Rap1p with N-terminal deletions. These results show that the N-terminal domain is not, in fact, dispensable for function. Since the NTD is resistant to further proteolysis once it is cleaved from the DNA binding domain, we believe that it forms a stable structural domain of its own. The ability to express and purify the isolated N-terminal domain supports this hypothesis.

Since Rap1p, Rap1 NTD-DBD, Rap1 DBDTEL, and Rap1 DBD all bind DNA with similar low picomolar dissociation constants, it appears that the DNA binding domain may well provide essentially all of the important interactions with respect to sequence-specific DNA binding. This result is consistent with results from other laboratories and with the recently determined structure of the Rap1 DNA binding domain complexed with DNA (Konig et al., 1996).

We found that Rap1p diluted into certain solutions bound to DNA with reduced affinity (see Figure 2-10). Through systematic comparison and variation of dilution buffers, we found that including 100 µg/ml BSA in the dilution buffer is critical for maintaining the DNA binding affinity of Rap1p at low protein concentrations. Several possible explanations exist for this observation. Perhaps Rap1p denatures at low protein concentrations or perhaps it binds to the eppendorf tubes, pipet tips, etc. BSA could counteract either of these factors.

With these methods for expressing and purifying Rap1p as well as several structural domains of Rap1p, we were now able to proceed to crystallization and further

biochemical studies.

Chapter III

Cooperative Binding of Rap1p to Adjacent DNA Binding Sites Depends on Spacing Between Sites

Introduction

Cooperative interactions between proteins that bind DNA are of fundamental importance in biological systems. For example, cooperative binding and assembly of multiple components to promoter and enhancer elements is critical for proper regulation of transcription (Senear and Brenowitz, 1991). Coliphage HK022 repressor (Carlson and Little, 1993), *lac* repressor (Brenowitz et al., 1991a, Brenowitz et al., 1991b), the bacteriophage λ cI repressor/operator system (Senear et al., 1986); and the Hox gene products (transcription factors that are major determinants of embryonic patterning) (Shanmugan et al.,

1997) are just a few examples of transcriptional regulatory systems where cooperative DNA binding has been extensively studied. Cooperativity between stably-bound proteins is proposed to be the source of transcriptional regulation by yeast Heat Shock Factor and phosphorylation of this protein appears to modulate cooperativity (Sorger and Pelham, 1988). In many cases cooperative interactions yield important higher order nucleoprotein structures. For example, *lac* repressor and AraC protein can both form higher order "looped" DNA structures when operator sites are properly spaced (Kramer et al., 1987; Schleif, 1992). Interaction between *E. coli* Hu proteins produces a condensed, coiled DNA structure (Tanaka et al., 1984). Higher order structures arising from assemblies of histones have been very extensively studied and characterized.

Rap1p binding sites occur in tandem arrays in the irregular $(C_{1-3}A)_n$ repeat of *Saccharomyces cerevisiae* telomeres (Buchman et al., 1988; Longtine et al., 1989; Wang and Zakian, 1990). Analysis of this telomere sequence revealed that a perfect Rap1p consensus binding site occurs ~every 37 base pairs (Wang and Zakian, 1990);

however, results of *in vitro* footprinting experiments have lead to the proposal that Rap1p binds to telomeric DNA as often as every 18 base pairs (Gilson et al., 1993). Given multiple, closely-spaced Rap1p binding sites in yeast telomeres, the abundance of Rap1p at telomeres, and the existence of a non-nucleosomal higher order structure at yeast telomeres, might Rap1p itself provide the basis for a regular, higher order structure at yeast telomeres?

Our efforts to explore cooperativity in binding of Rap1p to adjacent DNA sites in order to understand how Rap1p-Rap1p interactions might provide for a higher order nucleoprotein structure at yeast telomeres will be described in this chapter. To address this question, we designed a set of DNAs containing two Rap1p binding sites separated by 0, 2, 4, 6, 8, 10, 12, and 16 base pairs and evaluated for cooperativity in binding the binding of Rap1p, Rap1 DBD, and Rap1 DBDTEL to these DNAs.

Materials and Methods

Binding Experiments

The following DNAs were synthesized on an Applied Biosystems 380B DNA synthesizer and purified by preparative polyacrylamide gel electrophoresis.

RBS AATTCACACC**ACACCCACACACCACACCCACACACC**ACACCG

RBS+2 AATTCACACC**ACACCCACACACC**GT**ACACCCACACACC**ACACCG

RBS+4 AATTCACACC**ACACCCACACACC**GTGT**ACACCCACACACC**ACACCG

RBS+6 AATTCACACC**ACACCCACACACC**GTGTGT**ACACCCACACACC**ACACCG

RBS+8 AATTCACACC**ACACCCACACACC**GTGTGTGT**ACACCCACACACC**ACACCG

RBS+10 AATTCACACC**ACACCCACACACC**GTGTGTGTGT**ACACCCACACACC**ACACCG

RBS+12 AATTCACACC**ACACCCACACACC**GTGTGTGTGTGT**ACACCCACACACC**ACACCG

RBS+16 AATTCACACC**ACACCCACACACC**GTGTGTGTGTGTGTGT**ACACCCACACACC**ACACCG

Both strands of each DNA duplex were 5' end-labeled using T4 polynucleotide kinase and γ^{32}P-ATP. The strands were mixed and then annealed by heating to 100°C and slow-cooling to room temperature.

In our first experiments using these DNAs we used sizing column buffer (5 mM Tris (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 0.1 mM EDTA, 2 mM DTT) as the protein dilution buffer and binding reactions were incubated for

30 minutes. We then realized that for optimal Rap1p binding to DNA, it is necessary to include BSA in the protein dilution buffer. We also determined that the Rap1p-DNA binding reaction requires 24 hours to reach equilibrium. Therefore, we attempted to repeat earlier experiments using the newly determined optimal binding conditions. Data collected from experiments using Rap1p were usable. However, technical problems in quantifying data collected from experiments with Rap1 DBD and Rap1 DBDTEL resulted in too much variability among experiments for these data to be useful. Therefore, experiments described here were conducted using two different binding conditions.

The initial experiments were conducted as follows (referred to as Binding Condition 1). Protein was diluted to 10x stocks in 5 mM Tris (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 0.1 mM EDTA, 2 mM DTT. These were added to 50 pM DNA, 50 mM Tris (pH 7.5), 80 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 5.0% glycerol, and 100 µg/ mL BSA, and incubated at room temperature for 30 minutes. These samples (5 µl) were loaded onto native polyacrylamide gels (7%, 99:1 acrylamide:bis for Rap1p or 10%, 55:1

acrylamide:bis; 1x TBE) at 4°C, and gels were run at 10 mA for 5 hours.

The second set of binding reactions was performed subsequently using the following procedure (referred to as Binding Condition 2). Protein was diluted to 2x stocks in 50 mM Hepes (pH 7.5), 80 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 5.0% glycerol, 100 µg/ mL BSA. It was then mixed 1:1 with 2 pM DNA in the same buffer. These reactions were incubated at room temperature for twenty-four hours. The samples (10 µl) were loaded onto a native polyacrylamide gel (7%, 99:1 acrylamide:bis; 1x TBE) at room temperature and run at 10 mA for 5 hours. Gels were dried, scanned using a PhosphorImager (Molecular Dynamics), and quantified using ImageQuant (Molecular Dynamics).

Data Analysis

Hill coefficients were calculated in KaleidaGraph using Equation 3.1:

Fraction Rap1p binding sites occupied=$(M0^{M1})/(M2+(M0^{M1}))$

M0=free protein concentration

M1=Hill coefficient

M2=K_d

Data collected from experiments using binding condition 2 were also fit to Model I (Senear and Brenowitz, 1991) using a program written by Martin Horvath.

Model I

$$\text{Rap1p+DNA} \underset{}{\overset{2k_1}{\rightleftharpoons}} \text{Rap1p:DNA}$$

$$\text{Rap1p+Rap1p+DNA} \underset{}{\overset{k_1k_{12}/2}{\rightleftharpoons}} \text{Rap1p:DNA:Rap1p}$$

k_1=binding constant for single site binding

k_{12}=cooperativity parameter

Rap1p is treated as a pre-existing dimer. The model assumes that the two binding sites are identical. K_{12}, the cooperativity parameter, reflects the interaction of molecules bound to adjacent sites.

Any site heterogeneity is incorporated into k_{12}. k_{12} <1 indicates negative cooperativity, k_{12} =1 indicates independent binding to the two sites, and k_{12} >1 indicates positive cooperativity. The curves were fit using a program in which k_1 and k_{12} were adjusted to minimize χ^2 values by a non-linear least-squares method of parameter estimation (Marquardt, 1963). Errors were estimated using the Monte Carlo "bootstrap" method (Efron, 1992). Data points were sampled from the original data with replacement to give 30 data sets. For each parameter, errors are one sample standard deviation observed in the analysis of these 30 resampled data sets.

Results

An example of a gel shift experiment in which Rap1p was bound to RBS+8 is shown in Figure 3-1. Unbound DNA, singly bound DNA, and doubly bound DNA are clearly resolved, making quantifying these species very straightforward.

Data from Rap1p binding experiments performed using binding condition 1 were fit to Equation 3-1, yielding

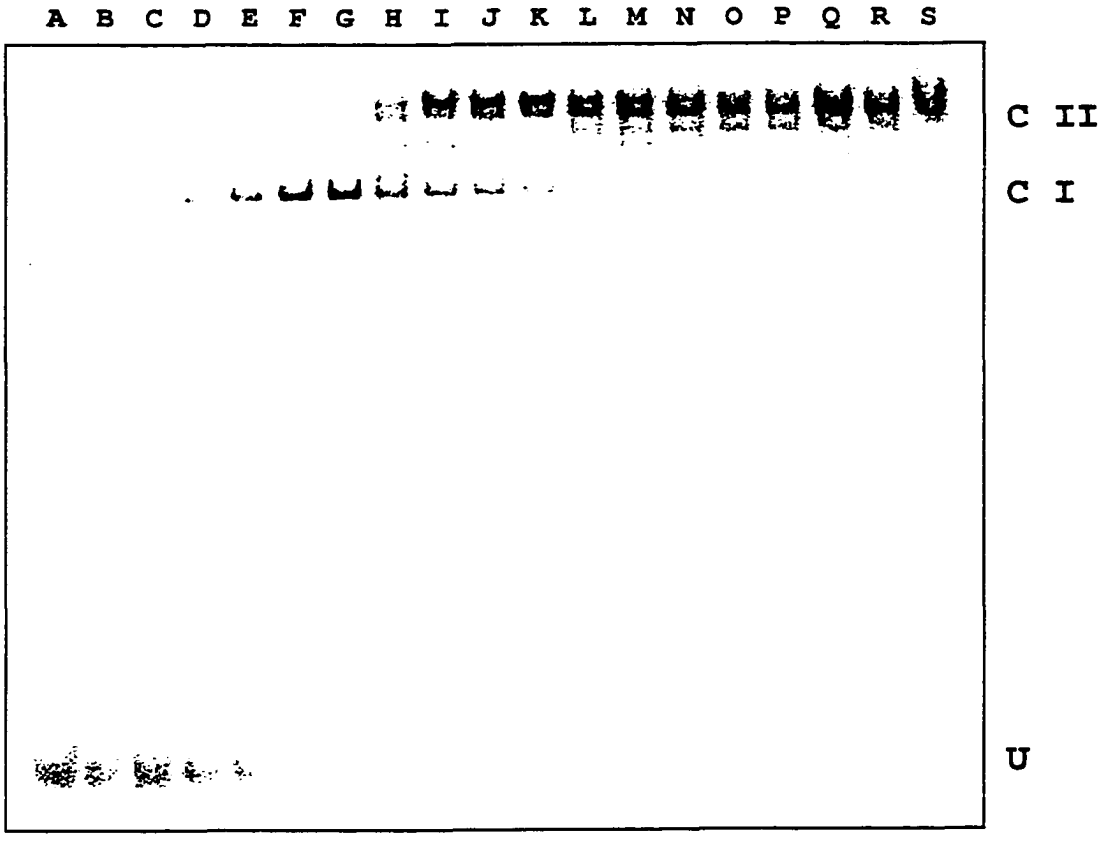


Figure 3-1 Rap1p Bound to RBS+8 Shown here is a native polyacrylamide gel used to quantify the binding of Rap1p to RBS+8. Unbound DNA (U), singly bound DNA (C I), and doubly bound DNA (C II) are clearly resolved. Protein concentrations (X 10^{-9} M) used in this experiment are as follows.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
0	.002	.004	.008	.02	.03	.06	.13	.25	.5	1	2	4	8	16	32	65	130	260

Hill coefficients which are listed in Table 3.1 A. Hill coefficients of ~1 indicate that Rap1p binds to two adjacent sites independently when the binding sites are separated by 0-10 base pairs.

This analysis was repeated and extended to include DNAs with two Rap1p binding sites separated by 12 and 16 base pairs, using Binding Condition 2. Hill coefficients determined from these experiments are reported in Table 3.1 B. These indicate that Rap1p binds independently to two adjacent Rap1p binding sites separated by 0-10 base pairs (Hill coefficients~1). However, Rap1p binds with positive cooperativity to two adjacent Rap1p binding sites separated by 12 base pairs (Hill coefficient=2) and Rap1p binds independently to two adjacent Rap1p binding sites separated by 16 base pairs (Hill coefficients ~1).

Data collected from experiments on Rap1p binding to DNAs with two Rap1p binding sites separated by 0-16 base pairs were also fit to Model I. Cooperativity parameters (k_{12}) are reported in Table 3.1 B. They suggest that Rap1p binds independently to adjacent binding sites separated by 0-8 base pairs (k_{12}~1), with increasing

Table 3.1 Hill Coefficients (HC) and Dissociation Constants (k_d) for Rap1p Binding to DNAs Containing Two Adjacent Binding Sites Separated by 2-10 Base Pairs (RBS+ 2-10). (A) Shown are HCs and k_ds for Rap1p binding using Binding Condition 1. Rap1p binds independently to DNAs with 2 Rap1p binding sites separated by 0-10 base pairs (HC~1). (B) Shown are HCs, k_ds, k_{12}s and standard deviation of k_{12} for Rap1p binding using Binding Condition 2. Rap1p binds independently to DNAs with 2 Rap1p binding sites separated by <12< base pairs (HC~1, k_{12}~1). Rap1p binds with positive cooperativity to RBS+12 (HC=2, k_{12}=2).

A								
RBS+	0	2	4	6	8	10	12	16
HC	1.1	.8	1.2	1.0	1.0	1.0		
K_d	.8 nM	.3 nM	.9 nm	.1 nM	.9 nM	.7 nM		
B								
HC	1.0	0.7	0.9	.7	0.8	1.2	**2.0**	.9
k_d	30 pM	100 pM	70 pM	40 pM	50 pM	12 pM	20 pM	40 pM
k_{12}	0.7	0.5	0.7	0.5	0.7	1.4	**2.0**	0.6
k_{12} SD	0.1	0.1	0.1	0.1	0.1	0.3	0.4	0.1

cooperativity to adjacent binding sites separated by 10 and 12 base pairs (k_{12}= 1.4 and 2.0, respectively), and independently again to adjacent binding sites separated by 16 base pairs (k_{12}~1).

K_ds determined using Binding Condition 1 range from 100pM to 900 pM. k_ds determined using Binding Condition 2 are approximately 1 order of magnitude tighter than those determined using Binding Condition 1 (ranging from 12 pM to 100 pM) as expected because Binding Condition 2 was optimized for tighter binding.

Data from Rap1 DBDTEL binding experiments performed using binding condition 1 were fit to Equation 3-1, yielding Hill coefficients which are listed in Table 3.2. Rap1 DBDTEL binds to two adjacent Rap1p binding sites independently when the binding sites are separated by 2-6 base pairs (Hill coefficient~1), with positive cooperativity when the binding sites are separated by 8 base pairs (Hill coefficient=1.7), and apparently returns to binding independently when the binding sites are separated by 10 base pairs (Hill coefficient=1.3).

Data from Rap1 DBD binding experiments performed using binding condition 1 were fit to Equation 3-1,

Table 3.2 Hill Coefficients (HC) and Dissociation Constants (k_d) for Rap1 DBDTEL Binding to DNAs Containing Two Adjacent Binding Sites Separated by 2-10 Base Pairs (RBS+ 2-10). Rap1 DBDTEL binds with positive cooperativity to RBS+8 which contains 2 Rap1p binding sites separated by 8 base pairs (HC=1.7) and independently to two adjacent binding sites separated by 2, 4, 6 or 10 base pairs (HC~1). K_ds range from 2-9 nM.

RBS +	2	4	6	8	10
HC	1.1	0.8	1.0	**1.7**	1.3
k_d	2.0 nM	8.0 nM	4.0 nM	3.0 nM	9.0 nM

yielding Hill coefficients which are listed in Table 3.3. Rap1 DBD binds to two adjacent Rap1p binding sites independently when the binding sites are separated by 0, 4, or 8 base pairs (Hill coefficient~1).

Discussion

Because Rap1p binding sites occur in tandem arrays in yeast telomeric DNA, we wished to determine whether cooperative interactions between Rap1p molecules exist and whether such interactions might play a role in assembling a higher order structure at yeast telomeres. To accomplish this we constructed a series of DNAs containing two Rap1p binding sites separated by 0-16 base pairs for use in binding experiments with Rap1 DBD, Rap1 DBDTEL and Rap1p. We found that Rap1 DBD binds independently to all sites tested, while Rap1 DBDTEL binds independently to two adjacent Rap1p binding sites separated by less than or more than 8 base pairs and with positive cooperativity to two adjacent Rap1p binding sites separated by 8 base pairs. Full length Rap1p binds

Table 3.3 Hill Coefficients (HC) and Dissociation Constants (k_d) for Rap1 DBD Binding to DNAs Containing Two Adjacent Binding Sites Separated by 0, 4, or 8 Base Pairs (RBS+ 0, 4, 8). Rap1 DBD binds independently to two Rap1p binding sites separated by 0, 4, or 8 base pairs (HC~1). K_ds range from 300 pM to 1.6 nM.

RBS +	0	4	8
HC	1	.8	1.2
k_d	1.0 nM	0.3 nM	1.6 nM

independently to two adjacent Rap1p binding sites separated by less than or more than 12 base pairs and with positive cooperativity to two adjacent sites separated by 12 base pairs.

We have described experiments utilizing two different binding conditions. In the first set of experiments, we evaluated binding of Rap1p, Rap1 DBD, and Rap1 DBDTEL to DNAs containing two Rap1p sites separated by 0-10 base pairs using conditions for evaluating binding of Rap1p to DNA that were standard at that time. Subsequently we found that including BSA in the protein dilution buffer gave 5 to 10 fold tighter binding so we switched to a dilution buffer containing BSA for a second set of experiments (Binding Condition 2). In these experiments, we reevaluated binding of Rap1p to adjacent sites separated by 0, 2, 4, 6, 8, and 10 base pairs and expanded the study to include sites separated by 12 and 16 base pairs. These were difficult experiments in that the very low DNA concentration (1 pM) resulted in less accurate quantitation because of a poor signal:noise ratio and, therefore, data of only modest quality. Nevertheless, we obtained similar results from

experiments conducted using Binding Conditions 1 and 2.

In analyzing data collected from the latter experiments, we used the method described by Senear and Brenowitz (Senear and Brenowitz, 1991) which relies on measurement of unbound, intermediate and fully bound DNA for each protein concentration.

Rap1p binding sites occur ~every 35-40 base pairs (that is, at sites separated by 22-27 base pairs) in the irregular *S. cerevisiae* telomeric repeat. It has been reported, based on *in vitro* footprinting experiments, that Rap1p can bind to the telomeric repeat as often as every 18 base pairs (that is, at sites separated by 5 base pairs) (Gilson et al., 1993). This is remarkably similar to the optimal spacing of 12 base pairs between Rap1p binding sites that we have observed, particularly considering differences in materials and techniques that were used in the two experiments.

It is interesting that the telomeric DNA in *K. lactis* contains a regular repeat of 25 base pairs, which is the Rap1p binding site interval predicted by the experiments described in this chapter. *K. lactis* and *S. cerevisiae* are budding yeasts which share Rap1p homologs

(Larson et al., 1994). While the telomeric DNA sequence in *S. cerevisiae* is irregular repeat $(C_{1-3}A)_n$, the telomeric repeat of *K. lactis* is a regular repeat of 25 base pairs. It is possible that this interval will be relevant to understanding how frequently *S. cerevisiae* Rap1p binds DNA at telomeres.

Our working model is that Rap1 DBD can be physically accommodated by a 13 base pair DNA site, but no protein-protein interaction occurs between two proteins bound to adjacent Rap1p binding sites. In Rap1 DBDTEL binding, the C-terminal domain provides favorable protein-protein interactions when two Rap1p binding sites are separated by 8 base pairs, yielding an optimal site size of 21 base pairs. In Rap1p binding, protein-protein interactions provide for positive cooperativity when the sites are separated by 12 base pairs. The Hill coefficient of 2 for Rap1p bound to RBS+12 is similar to that observed for Rap1 DBDTEL bound to RBS+8 (1.7), suggesting that the C-terminal domain may be involved in the Rap1p-Rap1p interaction as in the case of Rap1 DBDTEL. It seems that there is a narrow window of spacing between Rap1p binding sites that promotes positively cooperative Rap1p-Rap1p or

Rap1 DBDTEL-Rap1 DBDTEL interactions. When spacing is less than or greater than the optimum, the proteins bind independently to the two binding sites.

While it is formally possible that the Rap1p dimer (which will be discussed in chapter 5) occupies both binding sites on one apparently singly-bound DNA, we do not believe this is the case. The binding affinity of Rap1p for the first site is comparable to the affinity of the Rap1 DBD, which is a monomer. It is also similar to the affinity of Rap1p for a DNA containing a single site. If a Rap1p dimer were binding both sites on one DNA, the binding affinity for the first site should be considerably higher than the binding affinity of a monomer (Rap1 DBD or Rap1 DBDTEL) for the first site.

The results of these experiments also provided new directions for our crystallization efforts. We incorporated the information on cooperative DNA binding into the design of oligonucleotides for use in our crystallization experiments. These experiments are described in chapter IV.

Chapter IV

Crystallization

Introduction

Although our crystallization efforts included full length Rap1p and the domains defined by proteolysis, our primary efforts focused on crystallization of Rap1p complexed with DNA, since structural information on the complex would be more informative and of greater interest. Crystallization of protein-DNA complexes is extremely dependent on DNA length, 5' and 3' terminal extensions, and DNA sequence (Schultz et al., 1990). In nearly all crystals of protein-DNA complexes, the DNA stacks end to end in a pseudo-continuous manner which, in part, defines the crystal lattice. Unfortunately, it is impossible to predict what DNA lengths and terminal extensions will yield crystals. Since Rap1p binds at multiple sites in telomeric DNA and appears to bend the DNA, we hoped to accommodate whatever higher order structure might exist in telomeres into our crystal

lattice. We first attempted to use short oligos that would anneal in a manner that would mimic yeast telomeres, but found that the G-rich DNA strand crystallized too readily on its own. Therefore, we turned to the cooperativity studies described in the previous chapter in attempting to define appropriate DNA sites.

For Rap1 DBDTEL, positive cooperativity was observed when Rap1p binding sites were separated by 8 base pairs. Therefore, we designed double-stranded DNAs for use in crystallization trials which would provide this spacing between Rap1p sites. Several of these oligonucleotides, when complexed with Rap1 DBDTEL did, indeed, yield crystals. Our initial analysis for full length Rap1p bound to DNA with two Rap1p sites separated by ≤10 base pairs indicated independent binding to the two sites. More recently, we have extended this analysis and found evidence of positive cooperativity when two Rap1p sites are separated by 12 base pairs. Crystallization trials of full length Rap1p with oligonucleotides that provide this spacing are planned.

During the course of this work, we learned that

Daniella Rhodes at the MRC had obtained crystals of the DNA binding domain of Rap1p complexed with DNA. Therefore, we concentrated our efforts on crystallization of full length Rap1p and Rap1 DBDTEL with DNA.

Materials and Methods

Crystallization trials

For the proteins alone, various precipitants (polyethylene glycol (PEG), ammonium sulfate), salts (NaCl, KCl, LiCl, $MgCl_2$), pH (5.0-8.0) and a variety of additives (2-methyl-2,4-pentanediol (MPD), ethylene glycol, b-octyl glucopyranoside, ethanol) were used in crystallization trials. For protein-nucleic acid complexes, the DNAs listed in Table 4.1 were used. For crystallization of the complexes, we focused primarily on low salt (<300 mM NaCl) conditions using PEG as the precipitant. Salts, pH, and additives were the primary variables in these crystallization efforts.

RAP1 DBDTEL complexed with DNA

Crystals of Rap1 DBDTEL complexed with DNA 8 were grown at 24°C using the hanging drop vapor diffusion

Table 4.1 DNAs used in crystallization trials DNA sequences that yielded crystals of protein-DNA complexes are indicated, with limit of diffraction where data were collected. Sequence printed in bold is the Rap1p consensus sequence.

	DNA Sequences	Crystals	Diffraction
1	**ACACCCAC ACACC**CAC **GTGGGTGT GTGGG**TGT	**YES**-G-rich DNA only	
2	CCAC**ACAC CCACACAC** **GGGTGTGT GGGTGTGT**	**YES**-G-rich DNA only	
3	**RB+6** CCC**ACACCCACACACC**CAC GGG**TGTGGGTGTGTGGG**TG		
4	**RB+7** ACCC**ACACCCACACACC**CAC TGGG**TGTGGGTGTGTGGG**TG	**YES**-complex with Rap1 DBDTEL	
5	**RB+8** ACCC**ACACCCACACACC**CACC TGGG**TGTGGGTGTGTGGG**TGG	**YES**-complex with Rap1 DBDTEL	
6	**RB+9** CACCC**ACACCCACACACC**CACC GTGGG**TGTGGGTGTGTGGG**TGG	**YES**-complex with Rap1 DBDTEL	
7	**RB+8, 3' overhang** CCC**ACACCCACACACC**CACCA TGGG**TGTGGGTGTGTGGG**TGG		
8	**RB+8, 5' overhang** CACCC**ACACCCACACACC**CAC TGGG**TGTGGGTGTGTGGG**TGG	**YES**-complex with Rap1 DBDTEL	**4.2Å**
9	**RB+8, 5' overhang 2** CCACCC**ACACCCACACACC**CA TGGG**TGTGGGTGTGTGGG**TGG	**YES**-complex with Rap1 DBD	**3.2Å**
10	**RB+9, 3' overhang** ACCC**ACACCCACACACC**CACCC GTGGG**TGTGGGTGTGTGGG**TGG		
11	**RB+9, 5' overhang** CCACCC**ACACCCACACACC**CA GTGGG**TGTGGGTGTGTGGG**TGG	**YES**-complex with Rap1 DBDTEL	
12	**RB+9, 5' overhang 2** CCCACCC**ACACCCACACACC**CA GTGGG**TGTGGGTGTGTGGG**TGG		

method. Drops (4 µl) containing complex of 0.1 mM RAP1 DBDTEL and 0.1 mM DNA 8, 7.5 % PEG 4000, 5% MPD, 50 mM MES (pH 5.5), 100 mM NaCl, 2 mM DTT, and 0.02 % NaN_3 were equilibrated against 1 ml of reservoir buffer containing 15% PEG 4000, 10% MPD, 100 mM MES (pH 5.5), 100 mM NaCl, 2 mM DTT, and .02% NaN_3. Crystals appeared in approximately three days and grew over about five weeks to dimensions of approximately 100x100x30 mm. Larger crystals were grown by macroseeding methods as follows.The crystals were harvested into stabilizing buffer (20% PEG 4000, 10% MPD (pH 5.5), 200 mM NaCl, 2 mM DTT, and .02% NaN_3), and washed three times by transferring to fresh stabilizing buffer in harvesting dishes. The single crystal was then washed three times on a cover slip in 7.5 % PEG 4000, 5% MPD, 50 mM MES (pH 5.5), 100 mM NaCl, 2 mM DTT, and .02 % NaN_3. Finally, 10 µl of 0.1 mM RAP1 DBDTEL and 0.1 mM DNA 8, 7.5 % PEG 4000, 5% MPD, 50 mM MES (pH 5.5), 100 mM NaCl, 2 mM DTT, and .02 % NaN_3 were placed on the cover slip with the crystal and equilibrated against 1 ml of reservoir buffer containing 15% PEG 4000; 10% MPD, 100 mM MES (pH 5.5);

100 mM NaCl; 2 mM DTT; and .02% NaN_3. The seeded crystals continued to grow for approximately three weeks to dimensions of about 200x240x60 mm.

RAP1 DBD complexed with DNA

Crystals of RAP1 DBD complexed with DNA 9 were obtained using the hanging drop vapor diffusion method. Drops (4 µl) containing 0.1 mM RAP1 DBDTEL and 0.1 mM DNA 9, 5% PEG 4000, 50 mM MES (pH 6.5), 100 mM NaCl, 2 mM DTT, and .02% NaN_3 were equilibrated against 1 ml of reservoir buffer containing 10% PEG 4000, 100 mM MES (pH 6.5), 100 mM NaCl, 2 mM DTT, and .02% NaN_3. Crystals appeared in 7-9 days and continued to grow over three weeks to dimensions of approximately 80x80x400 mm.

RAP1 N-terminal domain

Crystals of RAP1 NTD were grown at 24°C using the hanging drop vapor diffusion method. Drops (4 µl) containing 0.13 mM protein, 7.5 % PEG 4000, 50 mM Tris (pH 7.0), 100 mM NaCl, 2 mM DTT, and .02 % NaN_3 were equilibrated against 1 ml of reservoir buffer containing 15% PEG 4000, 100 mM TRIS (pH 7.0), 100 mM NaCl, 2 mM

DTT, and .02% NaN_3. Needle-shaped crystals grew over a period of four to six weeks.

Diffraction analysis

The crystals were mounted using conventional methods at room temperature or were flash-frozen in liquid propane. Data were collected on a Rigaku RAXIS IIC image plate system mounted on a Rigaku RU200 rotating anode x-ray generator equipped with the MSC/Yale-designed focusing mirrors. The crystals were maintained at 24°C, -20°C or -170°C (for flash-frozen crystals) during data collection. Data were analyzed using MSC software and Denzo.

Results

Rap1 DBDTEL complexed with DNA 8

Crystals of form I (Figure 4-1) initially grew to 60x60x20 mm. After macroseeding, they grew to 200x240x60 mm. These crystals are cubic, space group I432, have unit cell dimensions a=b=c=235Å, $\alpha=\beta=\gamma=90^{\circ}$, and diffract to 4.2 Å resolution. Neither cooling crystals during

Figure 4.1 Cubic crystals of Rap1 DNA binding plus C-terminal domains complexed with a 20 base pair DNA. These crystals diffract to 4.2 Å resolution.

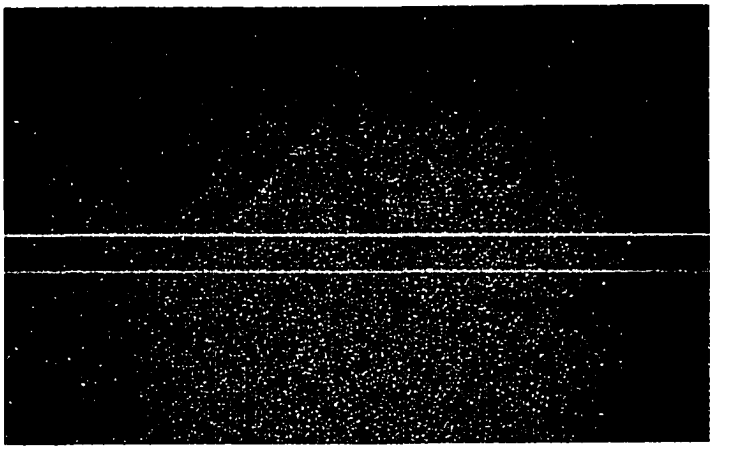


Figure 4.2 Crystals of Rap1 DNA binding domain complexed with a 19 base pair DNA. They diffract to 3.2 Å resolution.

data collection nor flash freezing improved the limit of resolution.

Rap1 DBD complexed with DNA 9

Crystal form II (Figure 4-2) contained Rap1 DBD complexed with DNA 9, though the drops were set up with Rap1 DBDTEL. Rap1 DBDTEL was apparently proteolyzed to Rap1 DBD when complexed with DNA 9, indicating that the protein is either stabilized when bound to DNA 8 or destabilized when bound to DNA 9. Crystals of form II were 80x80x400 mm rods, which diffracted to 3.2 Å resolution. These crystals are orthorhombic, space group P222, with unit cell dimension(a=71.9Å, b=94.9Å, c=106Å; $\alpha=\beta=\gamma=90°$.

Rap1 N-Terminal Domain

I also obtained crystals of Rap1 NTD. These grew as birefringent needles and needle clusters, but were not large enough for data collection.

Discussion

We began our efforts at crystallizing Rap1p with the goal of learning about the structure of Rap1p, its interactions with its cognate DNA, and potentially higher order structures of yeast telomeres. To this end, we embarked on crystallization trials of Rap1p and each of the structural domains that we had overexpressed and purified at that time. We obtained crystals of Rap1 N-terminal domain, which were promising, but not suitable for data collection because of their size.

It seemed likely that flexibility and poor stability of the unbound protein was contributing to problems with crystallizing Rap1p. In the hope that DNA binding would stabilize the protein as well as offer information about Rap1p's mode of binding, we designed two double stranded oligonucleotides for use in crystallization (DNAs 1 and 2 , Table 4.1). These contained 6 base pairs with a 5' or 3' one base overhang. We hoped that these DNAs would assume a telomere-like structure during crystallization with

Rap1p. However, we found that the G-rich strands had such a high propensity to crystallize alone that these DNAs were not useful in obtaining crystals of Rap1p complexed with DNA.

At this point it seemed important to determine what length of DNA would provide for optimal protein-protein interaction and for the formation of a pseudo-helical stacking of the DNA in the crystal lattice. To gather this information, we pursued the experiments described in Chapter III. We observed no cooperativity in the binding of Rap1p to the DNAs used in the initial experiments, but observed positive cooperativity in binding Rap DBDTEL to DNA that contained two binding sites separated by 8 base pairs. Therefore, we pursued crystallization of the Rap1 DBDTEL construct with a variety of DNAs suggested by our experimental results.

This lead to the successful crystallization of Rap1 DBDTEL with several DNAs (see Table 4.1), but none of the crystals were large enough for data collection. We expanded the set of DNAs to include

oligonucleotides with one or two base 5' or 3' overhangs and obtained crystals suitable for data collection from two of these DNAs: one complexed with Rap1 DBDTEL and one complexed with Rap1 DBD.

In crystal form I (Figure 4-1), Rap1 DBDTEL is complexed with DNA 8, which contains a Rap1p binding site in a 20 base pair oligo with a one base 5' overhang. These crystals grew to approximately 100x100x30 mm over a period of three weeks. By using these crystals as seeds for larger crystals, we obtained crystals that grew to approximately 200x240x60 mm. These crystals were cubic, space group I432, and diffracted to 4.2 Å resolution. However, I432 is a difficult space group and at a resolution of 4.2 Å, these crystals would not offer the detail we desired with regard to the structure of Rap1 DBDTEL complexed with DNA.

Crystal form II (Figure 4.2) was obtained from drops containing Rap1 DBDTEL and DNA 9, a 19 base pair oligonucleotide with a 2 base 5' overhang. These crystals grew as rods to approximately

80x80x400 mm over a period of about three weeks and were orthorhombic, space group P222. We were surprised to find that the only protein present in these crystals was Rap1 DBD. When bound to DNA 9 under crystallization conditions, Rap1 DBDTEL apparently underwent fairly rapid proteolytic cleavage to yield Rap1 DBD.

We considered crystallization trials with DNA 9 complexed with purified Rap1 DBD, but abandoned the idea when we learned that Daniella Rhodes of the MRC had obtained heavy atom derivatives and was well on the way to solving the structure of the Rap1 DNA binding domain complexed with DNA.

Since we now had data suggesting that the N-terminal domain, for which no clear function had been described, was mediating dimerization and tetramerization, we decided to pursue this finding using various biophysical and biochemical approaches. These experiments have now provided us with new insights into Rap1p function and have opened up new directions for our crystallization efforts.

Chapter V

Rap1 N-Terminal Domain Mediates Oligomerization

Introduction

It is generally accepted that Rap1p exists and functions as a monomer. This is based on two experiments. In the first experiment, (Muller et al., 1994) full length Rap1p and Rap1p Δ44-279 were co-translated *in vitro*, then incubated with DNA and analyzed by native polyacrylamide gel electrophoresis. If Rap1p were a dimer, one would expect to observe three species: a Rap1p (full length) homodimer, a heterodimer (one subunit full length, one subunit Δ44-279), and a Rap1p Δ44-279 homodimer. In this experiment, only two bands were detected, indicating that Rap1p was a monomer. In the second experiment (Muller et al., 1994), Rap1p was bound simultaneously to two lengths of DNA: a 27mer and a 134mer, each containing one Rap1p binding site. The expectation is that if Rap1p were a dimer, it would be capable of binding two oligonucleotides and, therefore,

yield three species which would be separable by native gel electrophoresis. Again, only two shifted bands were observed, indicating that Rap1p was behaving as a monomer.

These experiments suffer several shortcomings. In the second experiment, it is possible that Rap1p would not readily bind two 134mers or a 134mer and a 27mer or even two 27mers simultaneously due to steric and/or energetic considerations. In addition, Rap1p bound to a 134mer and a 27mer might not be easily distinguishable from Rap1p bound only to a 134mer on a native gel. In the first experiment, only an N-terminal deletion was explored. If the N-terminal region were important for dimerization, this experiment would yield incomplete information. Another difficulty with these experiments is that they utilized Rap1p produced by *in vitro* translation in rabbit reticulicyte lysates which may not allow for proper oligomerization of Rap1p. Indeed, despite much effort, we were unable to generate full length Rap1p from *in vitro* translation that migrated properly in native gels. Rap1 DNA Binding domain and Rap1 DBDTEL produced by *in vitro* translation behaved as

expected, but full length Rap1p did not. Because of these experimental difficulties, we believe that these experiments do not eliminate the possibility that oligomerization might be important for the function of Rap1p.

Since Rap1p binding sites occur in tandem arrays in the double-stranded $(C_{1-3}A)_n$ telomeric repeat (Longtine et al., 1989), Rap1p-Rap1p interaction would potentially have very important implications with respect to telomere structure. Therefore, we have explored the issue of Rap1p-Rap1p interactions in detail using the purified proteins described in Chapter II. We present evidence here that Rap1p forms stable dimers and tetramers and that this dimerization/tetramerization is mediated by the N-terminal domain. Though the N-terminal domain of Rap1p has been reported to be necessary for DNA bending (Muller et al., 1994; Vignais et al., 1989), no specific function had previously been ascribed to the N-terminal domain.

MATERIALS and METHODS

Gel Filtration Column Experiments

Gel filtration experiments were conducted using a Biorad Bio-Gel SEC 40-XL HPLC Gel filtration Column (size range 10-1500 kD) or a Biorad Bio-Sil SEC-400-5 HPLC Gel filtration Column (size range 20-1000 kD) on a Waters HPLC system (Waters 510 HPLC Pumps and Waters 486 Tunable Absorbance Detector). The columns were equilibrated in buffer VI (5 mM Hepes (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 0.1 mM EDTA, and 2 mM DTT), and protein was injected onto the column in a volume of less than 200 µl. The proteins were eluted at a flow rate of 0.6 ml/min and peaks (280nM) were collected for gel analysis. BioRad gel filtration standards (catalog number 151-1901) were used to create the standard curve, which was obtained by plotting eluted volume/void volume vs. ln(molecular weight) for each protein standard. Void volume was measured as the eluted volume of blue dextran. Apparent molecular weights of Rap1p and its structural domains were obtained by using the equation of the line calculated from the standard curve. Errors were

estimated as the variance around the mean of eluted volume/void volume from at least three experiments.

Ferguson Gels

Transverse pore gradient gels were prepared as follows (Ferguson, 1964; Retamal and Babul, 1988). Solutions of 4.5%-12% (37.5:1 acrylamide:bis) were prepared in 375 mM Tris·HCl (pH 8.9), .29 µl/ml TEMED, 0.1% ammonium persulfate. Vertical sections of polyacrylamide (in 0.25% to 1% increments) of at least 4 cm were poured into gel plates separated by 0.75 mm spacers, allowing each section to polymerize before pouring the next. A 3.125% stacking gel (5:1 acrylamide:bis) in 62.5 mM Tris (pH 6.7), 0.1% ammonium persulfate, .58 µl/ml TEMED was poured horizontally over these sections. 100 pMol to 200 pMol protein was loaded onto each vertical segment of the gel in loading buffer (83 mM Tris (pH 6.7), 15% glycerol, 42 µg/ml Bromphenol Blue). The running buffer contained 5 mM Tris, 38 mM glycine, pH≈8.3. The gels were run at 125 volts until the dye front was near the bottom of the gel. The dye front was marked by cutting the gel along the dye front,

the gels were treated with coomassie blue (30% methanol; 10% acetic acid; .25% Coomassie Blue R) to stain the proteins, then washed with destain (30% methanol; 10% acetic acid) and dried.

Ferguson plots were constructed as follows (Ferguson, 1964; Retamal and Babul, 1988). Mobility relative to the dye front (R_f) was measured for each protein band at each acrylamide concentration. Standards used for these experiments were from Sigma with molecular weights ranging from 14.2 to 545 kD. For each protein, a Ferguson plot was constructed by plotting $100*\log(R_f x100)$ vs % acrylamide. The negative slopes were then plotted against molecular weight for each of the standard proteins. The molecular weights of Rap1p and its structural domains were calculated by using the equation of the line of the standard curve and the negative slope determined for each protein was used to solve for molecular weight. Errors were estimated as variance about the mean of the R_fs measured for at least three experiments for each protein.

For Ferguson analysis of full length Rap1p and structural domains bound to DNA, 100pM ^{32}P labeled DNA was

added to 1 nM protein in loading buffer (83 mM Tris (pH 6.7), 15% glycerol, 42 µg/ml Bromphenol Blue) and incubated at room temperature for 1 hour. 10 µl of each sample were loaded onto the gel and the gels were electrophoresed for 3 hours at 125 V. Gels were cut off at the dye front, dried, and the dye front marked with radioactive label. Gels were scanned using a PhosphorImager (Molecular Dynamics). Molecular weights were calculated and errors estimated as described for unbound proteins. DNA used for this experiment is as follows:

5'ACC**ACACCCACACACC**CAC 3'
TGG**TGTGGGTGTGTGG**GTG

Construction of plasmids encoding RAP1 N-terminal modifications

Expression clones encoding RAP1 with N-terminal extensions or deletions were constructed as follows. The gene encoding RAP1 with seven lysine residues added to the N-terminus was constructed by ligating the following oligonucleotide into the EcoRI site of pKKT7-RAP1 to create plasmid pKKT7-RAP1Nlys7. Plasmids containing the insert have a unique AseI site and no longer contain an

Eco RI site and were selected for by digestion with EcoRI prior to transformation. Expression of modified Rap1p was verified by western blotting.

NT3 Linker

```
            M   K   K   K   K   K   K   K   F   M   S
     AA TTA ATG AAA AAG AAG AAG AAA AAG A
          T TAC ATT TTC TTC TTC TTT TTC TTT AAg tac aga
```

Expression clones encoding RAP1 containing N-terminal deletions of 10 and 15 residues were generated by ligating the following three fragments together to create plasmids pKKT7-RAP1 Δ15 and pKKT7-RAP1Δ10:

Fragment 1 pKKT7-RAP1 5063 bp EcoRI-Blp1 fragment

Fragment 2 pKKT7-RAP1 784 bp BamHI-BlpI fragment

Fragment 3 synthesized Linker 1 or Linker 2.

Linker 1(Δ15)

```
                    M   A   L   D
              AA TTC ATG GCA TTG
                   G TAC CGT AAC CTA G
```

Linker 2 (Δ10)

```
                M   P   A   E   Y   V   D   A   L   D
          AA TTC ATG CCA GCA GAA TAT GTT GAT GCA TTG
               G TAC GGT CGT CTT ATA CAA CTA CGT AAC CTAG
```

Purification of Rap1p N-terminal Modifications

Plasmids pKKT7-RAP1Nlys7, pKKT7-RAP1Δ15 and pKKT7-RAP1Δ10 were transformed into *E. Coli* BL21(DE3) pLysS

cells by electroporation. Cells were grown at 37°C to OD_{600}=0.5, cooled to room temperature, and production of the Rap1p mutants induced by the addition of IPTG to a final concentration of 0.5 mM at OD_{600}=1.0. After growing at room temperature for an additional ~16 hours, cells were harvested by centrifugation and frozen at -15°C. Cells were resuspended in lysis buffer (50 mM Tris (pH 7.5), 50 mM NaCl, 0.02% NaN_3, 1mM EDTA, 2 mM DTT), lysed by sonication and cell debris was removed by centrifugation. Ammonium sulfate was added to 30% of saturation and precipitate removed by centrifugation. The N-terminal mutant forms of Rap1p were precipitated by addition of ammonium sulfate to 60% of saturation. The remainder of the purification follows the same protocol used for Rap1p (described in Chapter II). Yields were 1-2 mg from a 3 L cell growth.

Proteolysis of Rap1Δ10

As will be discussed in the results section, Rap1•10 is very resistant to proteolysis. Therefore, Rap1•10 was incubated in 2.0 M urea for 16 hours at 4°C, then subjected to proteolysis with trypsin (1:500,

protease:protein) for 30 minutes at 24°C. The course of the cleavage reaction was followed by SDS PAGE.

Rap1Δ10 Binding Experiments

DNA binding by Rap1Δ10 was evaluated as follows. Protein (untreated or treated in 2.0 M urea) was diluted into buffer VI (5 mM Hepes (pH 7.5), 0.2 M NaCl, 0.02% NaN_3, 0.1 mM EDTA 2 mM DTT) and then diluted 10X in binding buffer (50 mM Hepes (pH 7.5), 80 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 5.0% glycerol, 100 µg/ ml BSA) containing 1.0 pM ^{32}P-labeled DNA and incubated at room temperature for 1 hour. Bound DNA was separated from unbound by native polyacrylamide gel electrophoresis: 7%, 99:1 acrylamide:bisacrylamide, 1x TBE, 10 mA. Gels were dried and scanned using a PhosphorImager (Molecular Dynamics).

DNA binding by the proteolytic fragments of Rap1Δ10 was evaluated as follows. Protein from the proteolysis reaction was diluted 2x into binding buffer (100 mM Hepes (pH 7.5), 160 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 10.0% glycerol, 200 µg/ mL BSA) containing 100 pM DNA. This reaction was incubated at room temperature for one hour. The following DNA was used in these studies:

AATTCACACCC**ACACCCACACACC**ACACCACAG
GTGTGGG**TGTGGGTGTGTGG**TGTGGTGTCCTAG

RESULTS

Gel Fitration Analysis

Sample chromatographs of the gel filtration standards and of Rap1p are shown in Figure 5-1. A plot of elution volumes vs ln molecular weight is shown in Figure 5-2. The elution volumes and apparent molecular weights of Rap1p, Rap1 NTD+DBD, Rap1 DBD, Rap1 DBDTEL, and Rap1 NTD as determined by gel filtration analysis are listed in Table 5.1. Comparison of these values with the calculated molecular weights suggests that Rap1p, Rap1 NTD+DBD, and Rap1 NTD are tetrameric, and Rap1 DBD and Rap1 DBDTEL are monomeric in solution. Incubation of Rap1p with 2.0 M urea for 16 hours at 4°C does not alter the elution time indicating that treatment with 2 M urea is not sufficient to disrupt the oligomerization of Rap1p. Rap1Δ10, Rap1Δ15, and Rap1Nlys7 elute from the size exclusion column with an apparent molecular weight >>670 kD (see Figure 5-3). It appears that the N-

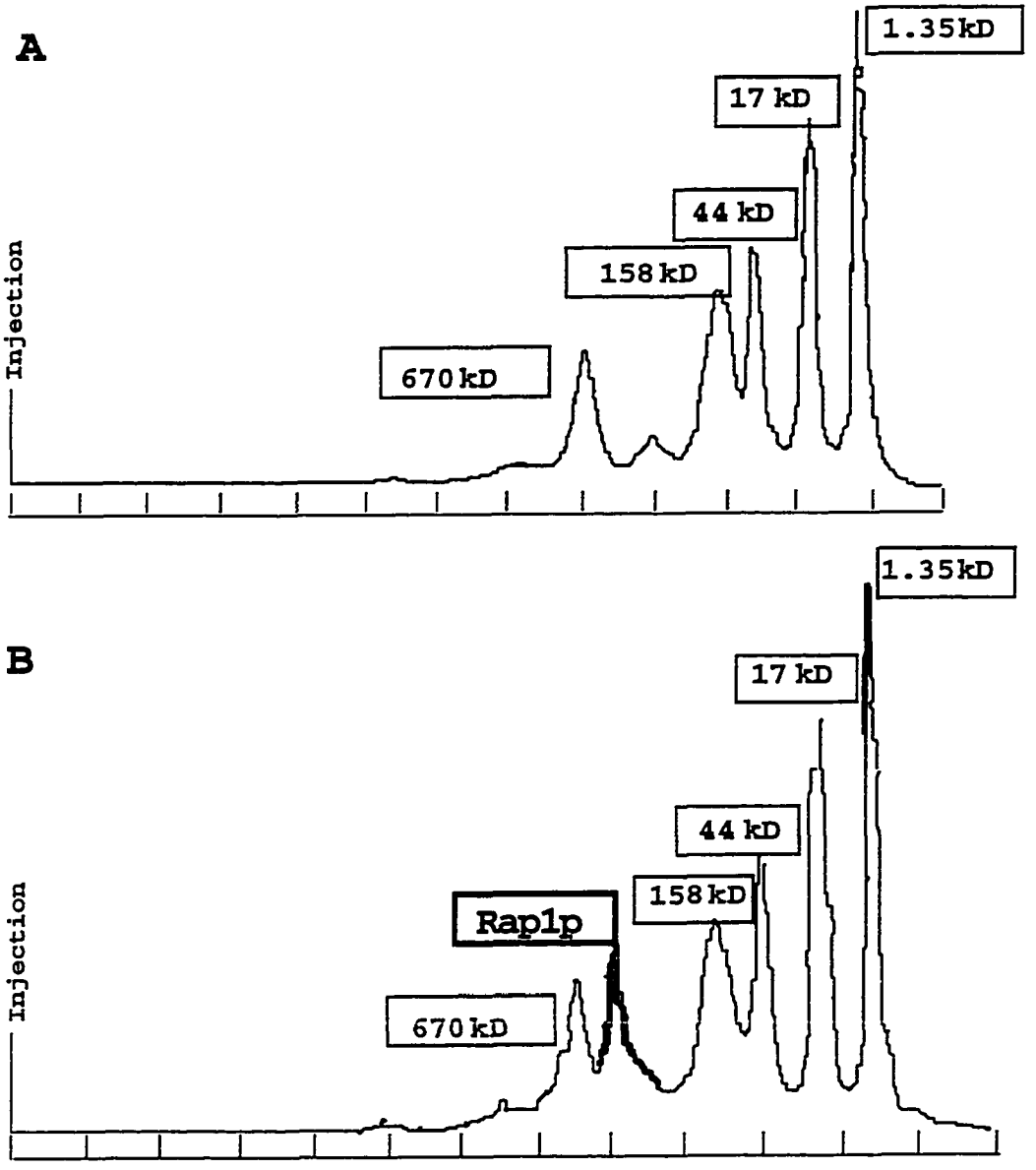


Figure 5-1 Chromatographs of Molecular Weight Standards and Rap1p Elution from HPLC Gel Filtration Column A. Elution profile of HPLC gel filtration column molecular weight standards. MW of each peak is indicated. **B.** Elution profile of molecular weight standards and Rap1p. Rap1p elutes with an apparent molecular weight of ~370 kD or four times its monomer size.

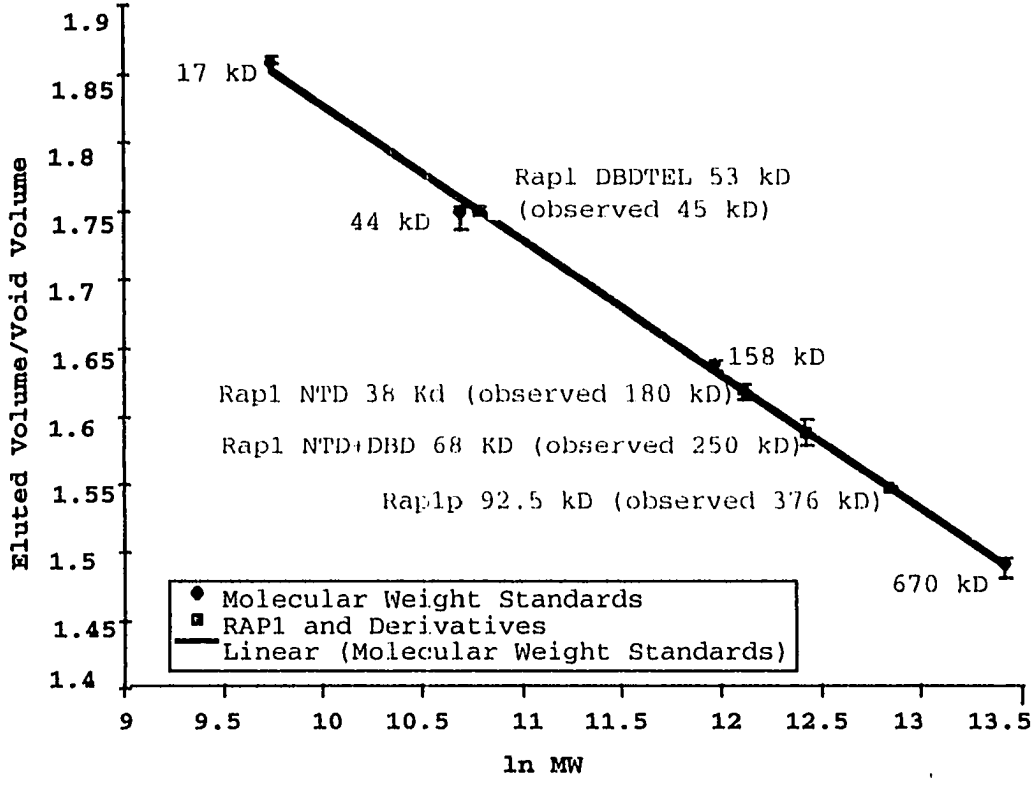


Figure 5-2 Gel Filtration Column Analysis of MW Standards and Rap1p Shown is a plot of elution volume/void volume vs ln(MW) from gel filtration column analysis of Molecular Weight Standards (blue diamonds). Apparent molecular weight of Rap1p and Structural domains based on elution volume from gel filtration column is indicated in pink.

Table 5.1 Listed here are elution volumes of molecular weight standards, Rap1p, and its structural domains from the gel filtration column. Apparent molecular weights of Rap1p and its structural domains were calculated from the standard curve. Subunits number is the apparent molecular weight divided by monomer molecular weight.

Protein	**Elution volume (ml)**	**Apparent MW (kD)**	**Monomer MW (kD)**	**subunits**
STANDARDS				
Thyroglobulin	7.5	670		
Bovine gamma globulin	8.2	158		
Chicken ovalbumin	8.8	44		
Equine myoglobin	9.3	17		
Rap1p and Structural Domains				
Rap1Δ10	6.1	>>670	91.4	??
Rap1Δ15	6.2	>>670	91	??
Rap1p	7.8	370	92	4
Rap1 NTD+DBD	8.0	250	67	4
Rap1 NTD	8.1	160	38	4
Rap1Δ10 after urea treatment	8.2	80	91.4	1
Rap1 DBDTEL	8.8	48	53	1
Rap1 DBD	9.4	20	28	1

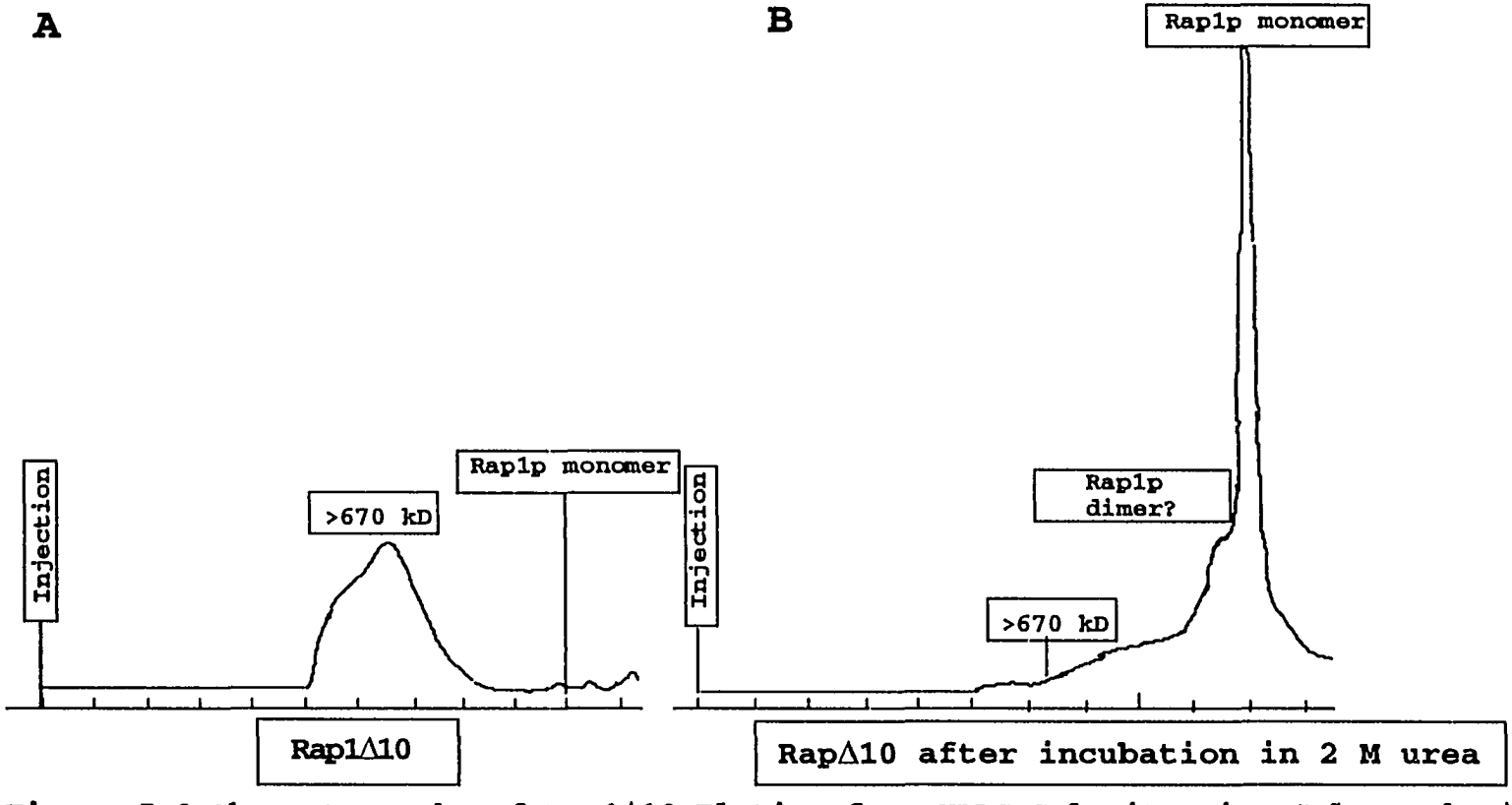


Figure 5-3 Chromatographs of Rap1Δ10 Elution from HPLC Gel Fitration Column Elution profiles of Rap1Δ10 from HPLC gel filtration column (A) before and (B) after treatment with 2 M urea. Before treatment with urea, Rap1Δ10 elutes as a large, soluble oligomer. After urea treatment, it elutes with the apparent molecular weight of a Rap1p monomer. There may be a very small amount of Rap1Δ10 dimer present after treatment with urea.

terminally modified Rap1 proteins form large, soluble oligomers. Treatment of Rap1Δ10 with 2.0 M urea for 16 hours disrupts the oligomer and the urea-treated Rap1Δ10 behaves as a monomer in our gel filtration system(see Figure 5-3). Treatment of Rap1Δ10 with 2.0 M urea for shorter times yields both monomer and aggregate, but no intermediate species are observed.

Ferguson Analysis

A plot of negative slopes vs molecular weights is shown in Figure 5-4. An example of a gel used for Ferguson analysis is shown in Figure 5-5. See Table 5.2 for a list of negative slopes measured for Rap1p and its structural domains and their apparent molecular weights as determined by Ferguson analysis. Comparison of these values to the calculated molecular weights indicate that Rap1p and Rap1 NTD exist as both dimer and tetramer in solution, Rap1 NTD+DBD exists as both dimer and monomer, and Rap1 DBDTEL is monomeric. When bound to DNA, Rap1p and Rap1 NTD+DBD appear to be dimeric while Rap1 DBD appears to bind DNA as a monomer. Ferguson analysis was not possible on Rap1Δ15 because of its extremely retarded gel mobility.

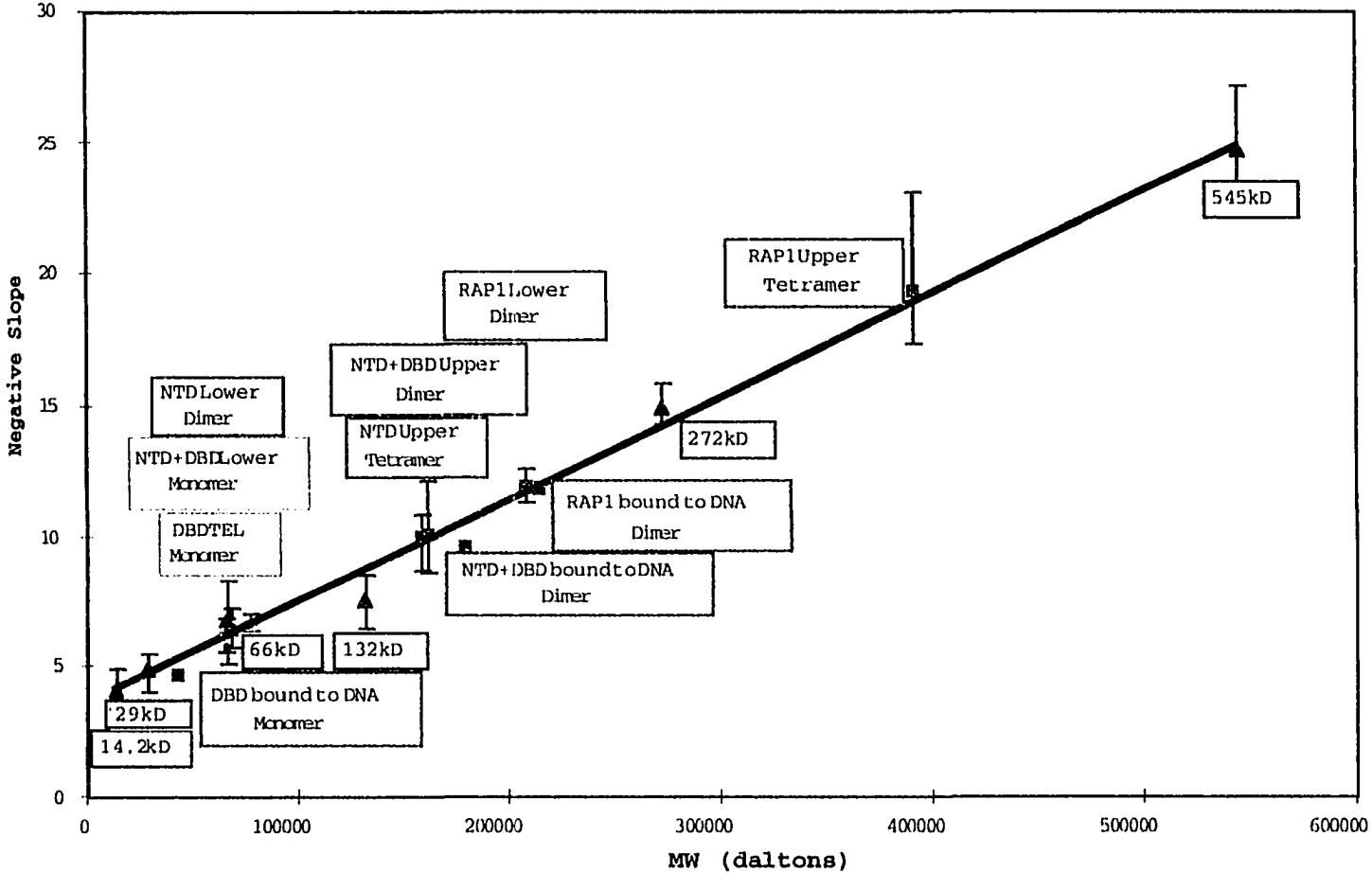


Figure 5-4 Ferguson Analysis of RAP1 and Derivatives Shown is a plot of negative slopes from Ferguson plots vs molecular weights of protein standards. Apparent oligomeric states of Rap1p and its structural domains are indicated in boxes (red=dimers and tetramers; yellow=monomers; blue=protein bound to DNA; green=molecular weight standards).

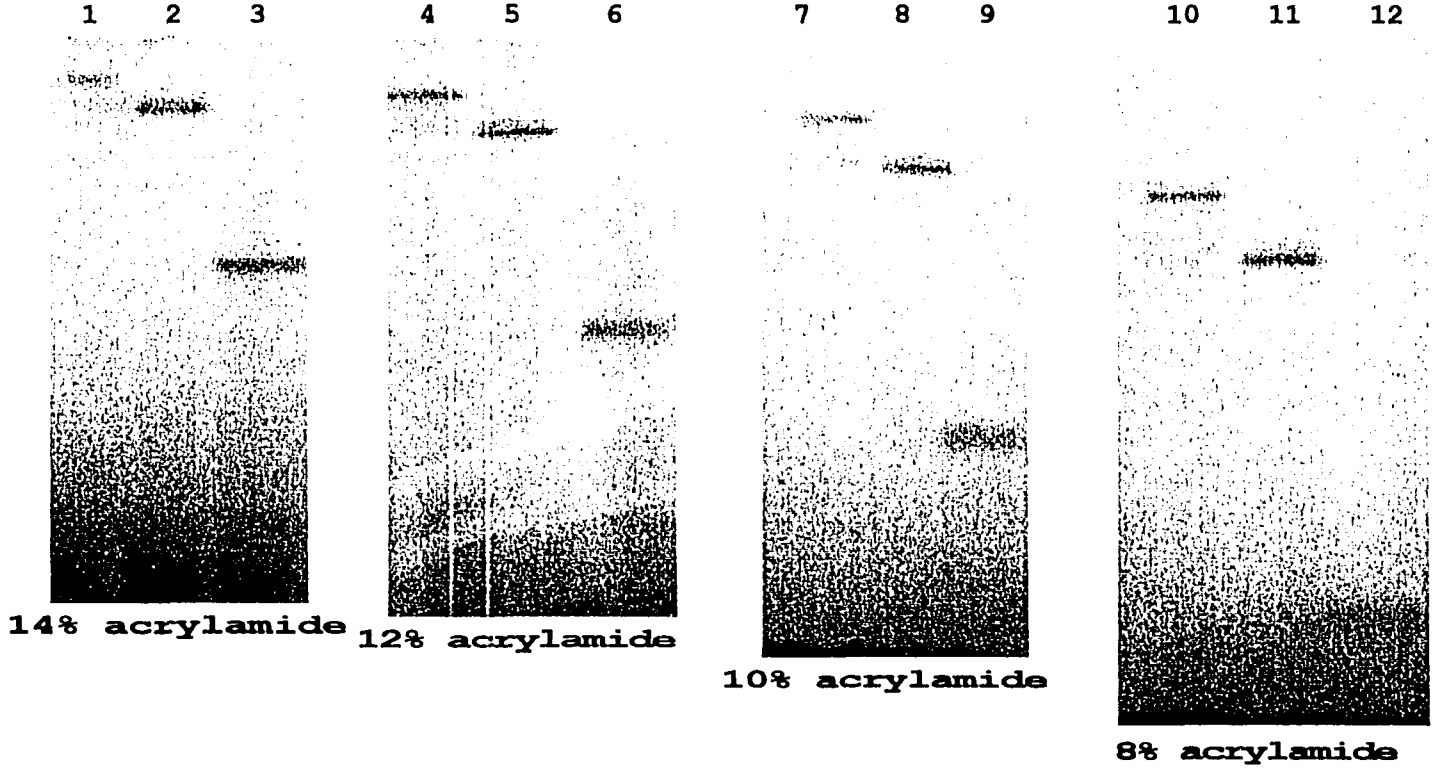


Figure 5-5 Examples of native gels used for Ferguson analysis. Rap1p (lanes 1,4,7), NTD+DBD (lanes 2,5,8), and DBD (lanes 3,6,9)bound to ^{32}P-labeled DNA have been run on polyacrylamide gels of four different acrylamide percentages. The relative rate of migration of each complex on the four gels depends on the molecular weight of the complex, with Rap1 being the largest and DBD the smallest.

Table 5-2 Listed here are negative slopes obtained from Ferguson analysis of protein migration for molecular weight standards, Rap1p, its structural domains and Rap1p complexed with DNA. The apparent molecular weight of Rap1p and its structural domains is calculated from the standard curve. Subunit number is the apparent molecular weight divided by the monomer molecular weight.

MW Standards	**Ferguson plot negative slope**	**Apparent MW (kD)**	**Monomer MW (kD)**	**Sub units**
α-Lactalbumin	4.0	14.2		
Carbonic Anhydrase	4.9	29		
BSA monomer	6.7	66		
BSA dimer	7.5	132		
urease trimer	15	272		
urease hexamer	24.7	545		
Rap1p and Structural domains				
Rap1p lower band	12	210	92.5	2
Rap1p upper band	19.3	390	92.5	4
Rap1 with DNA	12	210	107	2
NTD+DBD lower band	6.8	78	67	1
NTD+DBD upper band	10.0	160	67	2
NTD+DBD with DNA	9.1	140	81.5	2
NTD lower band	6.4	70	38	2
NTD upper band	10	160	38	4
DBDTEL	6.2	65	53	1
DBD with DNA	5.0	37	42.5	1

Rap1Δ10, Rap1Δ15, and Rap1Nlys7: DNA Binding and Proteolysis

Rap1Δ10, Rap1Δ15, and Rap1Nlys7 do not bind DNA using conditions that yield very tight binding by Rap1p. See Figure 5-6 for a comparison of DNA binding properties of Rap1Δ10 and Rap1p before and after incubation with 2M urea. In order to determine whether Rap1Δ10 was severely misfolded and, therefore no longer capable of binding DNA, we attempted to proteolyze Rap1Δ10. Rap1Δ10 was extremely resistant to proteolysis. Since gel filtration column analysis had indicated that treatment of Rap1Δ10 with 2.0 M urea disrupts the aggregated form to an apparently monomeric form, we tested the ability of urea-treated Rap1Δ10 to bind DNA. No binding was detected at protein concentrations below 1 μM (see Figure 5-6, lanes 20-25). In contrast, Rap1p retained normal DNA binding properties even after incubation in 2.0 M urea (see Figure 5-6, lanes 8-13) . At very high concentrations (5 μM), DNA binding can be detected for urea-treated Rap1Δ10, but not for untreated Rap1Δ10 (see Figure 5-7, lanes 6 and 4, respectively). Some DNA binding by

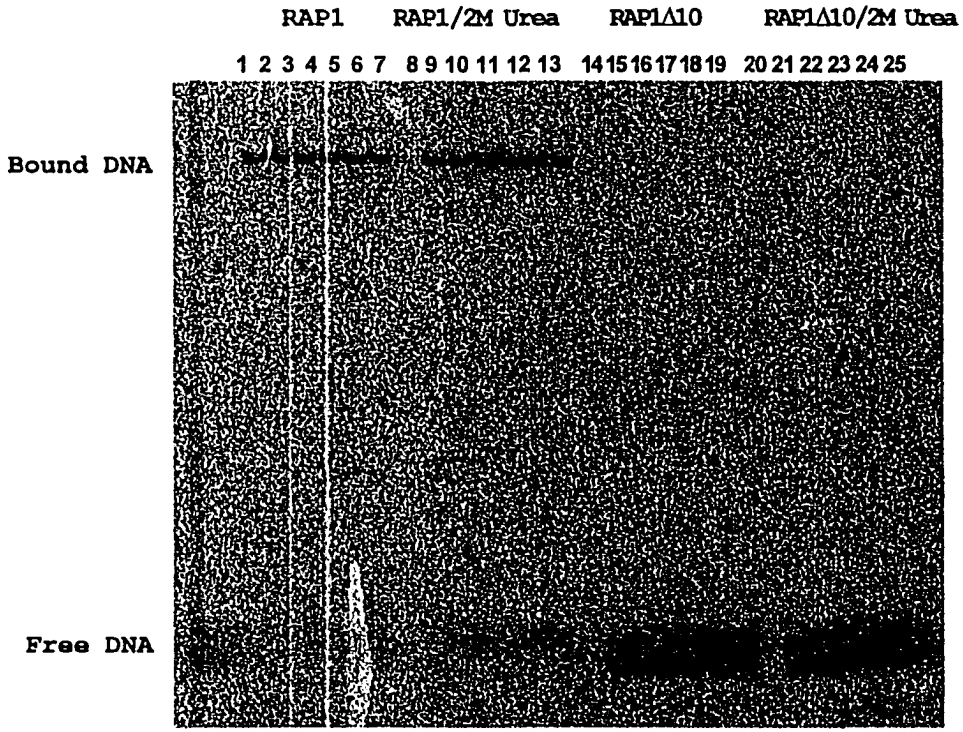


Figure 5-6 RAP1 and RAP1Δ10 +/- 2M Urea Gel Shift Experiment Shown is a native polyacrylamide gel comparing the DNA binding properties of Rap1p and Rap1Δ10. Rap1p binds DNA with high affinity before (lanes 2-7) and after incubation with 2M urea (lanes 8-13). Rap1Δ10 is not capabable of binding DNA either before (lanes 14-19) or after incubation with 2M urea (lanes 19-25). Protein concentrations (μM) are as follows: lane 1=0; lanes 2, 8, 14, 20=.001; lanes 3, 9, 15, 21=.016; lanes 4, 10, 16, 22=.039; lanes 5, 11, 17, 23=.063; lanes 6, 12, 18, 24=0.25; lanes 7, 13, 19, 25=1.0.

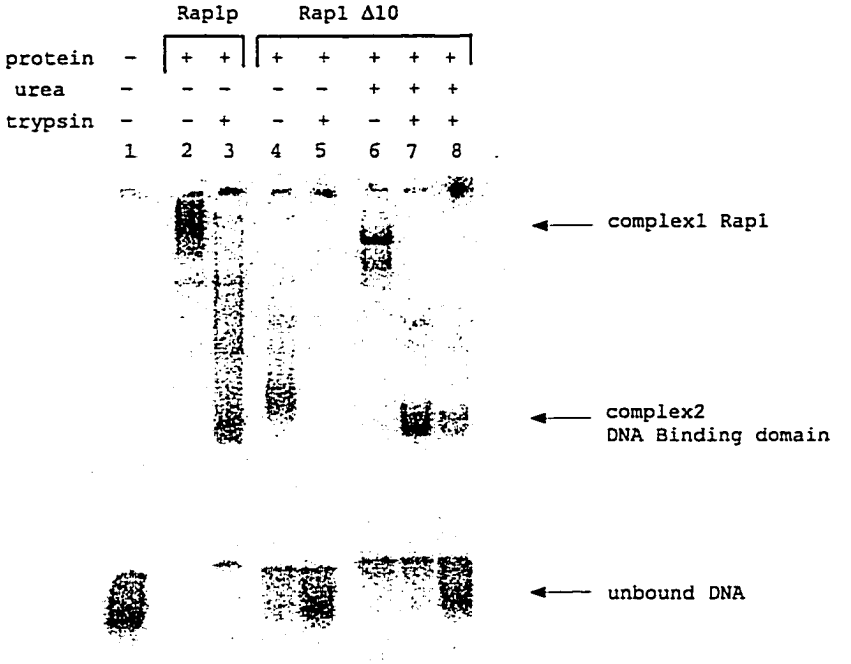


Figure 5-7 The effect of urea treatment and limited proteolysis on Rap1Δ10 DNA binding activity. Prior to treament with urea, Rap1Δ10 does not bind DNA even after limited proteolysis (lanes 4 and 5). After treatment with 2 M urea, Rap1Δ10 binds DNA at high protein concentration (lane 6) and limited proteolysis releases a DNA binding activity (lanes 7 and 8) that appears identical to that obtained after proteolysis of wild-type Rap1p (lane 3).

untreated Rap1Δ10 can be seen in Figure 5-7, lane 4, but since the mobility of the shifted DNA does not correspond to the mobility of bound Rap1p, we believe this is a non-specific binding activity. No DNA binding activity is released upon proteolysis of Rap1Δ10, (which is highly resistant to proteolysis, in contrast to Rap1p) However, Rap1Δ10 becomes more susceptible to proteolysis after treatment with 2.0 M urea and an apparently authentic DNA binding domain is released and can bind DNA (see Figure 5-7, lanes 7 and 8). These experiments show that removal of as few as ten residues from the N-terminus of Rap1p dramatically affects the structure of Rap1p, yielding a highly stable, but aggregated structure that no longer binds DNA and no longer forms wild-type dimers and tetramers.

DISCUSSION

We have presented evidence that Rap1p is at least a dimer in both its bound and unbound forms *in vitro*. Furthermore, by comparing the behavior of full length

Rap1p to that of modified forms from which the N-terminal domain, the C-terminal domain, or both have been deleted, we have shown that the N-terminal domain of the protein is a required mediator of this protein-protein interaction.

Size exclusion chromatography indicates that Rap1p, Rap1 NTD and Rap1 NTD+DBD are tetramers, while Rap1 DBD and Rap1 DBDTEL are monomers. Thus, Rap1p and structural domains that contain the N-terminal domain are tetramers and those that do not contain the N-terminal domain are monomers. The Rap1p tetramer is stable to Rap1p concentrations as low as 50 nM and to urea concentrations as high as 2.0 M, suggesting that the Rap1p molecules are not casually associated, but rather are quite tightly linked. The tetramer is competent to bind DNA as evidenced by the fact that when Rap1p is incubated with DNA, then passed over the gel filtration column, the DNA elutes with the Rap1p peak in a Rap1p-dependent fashion. These size exclusion column experiments suggest that Rap1p exists as a tetramer and that the N-terminal domain of Rap1p is required for oligomerization.

As an independent measure of the molecular weight of

Rap1p in solution, we utilized Ferguson analysis of the mobility of Rap1p in native polyacrylamide gels. Gel filtration measures Stokes' radius and, therefore, can be dependent on the shape of the molecule. Since we were not able to dissociate Rap1p into monomeric or dimeric species, we wanted to have an independent verification of the molecular weight. Electrophoretic mobility in polyacrylamide gels is dependent on the frictional coefficient as the molecule is pulled through the gel matrix. This analysis can also depend on molecular shape but in a different way than gel filtration analysis.

Ferguson analysis indicates that Rap1p and Rap1 NTD are in a dimer/tetramer equilibrium, Rap1 NTD+DBD is in a dimer/monomer equilibrium, and Rap1 DBDTEL is a monomer. Thus, again, Rap1p dimerization/tetramerization is dependent on the presence of the N-terminal domain. The presence of dimeric and tetrameric species for Rap1p and Rap1 NTD, and dimeric and monomeric species of Rap1 NTD+DBD permits even greater confidence that we are primarily measuring molecular weight and not molecular shape.

When Rap1p or Rap1 NTD+DBD bound to DNA was

subjected to Ferguson analysis, the complex had the apparent molecular weight of a dimer, while bound Rap1 DBD appears to be a monomer. We did not observe a tetramer band in the bound form of Rap1p that had been present in the unbound form, probably due to the fact that the experiments with protein bound to DNA utilized considerably lower protein concentrations than were used in the experiments with protein alone. The tetramer likely dissociates into dimer at the lower protein concentrations used in this analysis. For Rap1 NTD+DBD, substantial monomeric form would have been expected at the protein concentrations used in the DNA complex experiments. Only Rap1 NTD+DBD dimer was observed in this experiment, however, suggesting either that dimerization is required for DNA binding when the N-terminal domain is present, or that DNA binding favors dimerization.

The most direct way of determining the oligomeric state of a protein is to mix two forms of the protein and evaluate the presence of intermediate species. We tried a number of mixing experiments using the materials described in this thesis and have not observed any

intermediate species. These are summarized in Table 5.3. This may be due to an extremely slow exchange of subunits and/or it may be that oligomerization is highly sensitive to modifications throughout the protein (for example, Rap1p forms a more stable dimer/tetramer than NTD+DBD). Cotranslation *in vitro* proved unsuccessful because *in vitro*-translated Rap1p appeared not to be folded properly. We also attempted to bind different length DNAs to the Rap1p dimer/tetramer but these experiments suffered from difficulties in separating the different species and from our lack of knowledge regarding the relative dispositions of the DNA binding domains in the dimer/tetramer. To further strengthen our understanding of the oligomeric state of Rap1p, we set out to introduce modest modifications that might alter Rap1p-Rap1p interactions.

Since it appeared that the N-terminal domain of Rap1p is essential for oligomerization, we were interested in modifying the N-terminal domain in a manner that would disrupt Rap1p-Rap1p interaction. The published sequence comparison of Rap1p with its homolog from *K. lactis* reported near identity in the DNA binding

Table 5.3 Mixing Experiments
A. Purified Rap1p was mixed with rap1NTDp, rap1N-lys7p, or rap1C-lys7p and run on a native polyacrylamide gel. There was no indication of Rap1p oligomerization.
B. Rap1p was cotranslated *in vitro* with rap1NTD+DBDp or rap1NTDp in a rabbit reticulocyte lysate system and run on a native polyacrylamide gel. We were unable to obtain useful results because of problems with the gel mobility of Rap1p produced in this system.
C. Rap1p was incubated with these ^{32}P-labeled oligonucleotides which each contain one Rap1p binding site: a 33mer, a 334mer and a mixture of 33mer and 334mer. There was no indication of Rap1p oligomerization.

A

	Rap1NTDp	rap1N-lys7p	rap1C-lys7p
Rap1p	Purified proteins mixed	Purified proteins mixed	Purified proteins mixed

B

	rap1NTD+DBDp	rap1NTDp
Rap1p	*in vitro* co-translation	*in vitro* co-translation

C

	33mer	334mer	33mer+334mer
Rap1p	binding	binding	binding

domains, a high degree of conservation in the C-terminal domains, but complete divergence of the N-terminal domains. We noted, however, in comparing the N-terminal sequences of Rap1p from *S. cerevisiae* and the Rap1p homolog from *K. lactis*, that 7 of the first 10 amino acids are identical and one more is highly conserved (Larson et al., 1994). This high degree of sequence conservation suggests an important role for the extreme N-terminus of Rap1p and it seemed a reasonable guess that modification of this region might disrupt protein-protein interaction. Therefore, we mutated the N-terminus by adding 7 lysine residues (Rap1Nlys7) or deleting residues S2-A10 (Rap1Δ10) or S2-D16 (Rap1Δ15) (see Figure 5-8).

Although these proteins remained completely soluble, each appeared to form a very large aggregate (>>670 kD) as evaluated by gel filtration chromatography. This aggregate was easily disrupted by incubating the modified proteins in 2 M urea for sixteen hours at 4°C. The urea-disrupted protein now eluted from the size exclusion column with the apparent molecular weight of a monomer, our first observation of the monomeric form of Rap1p.

During the course of the disruption, both oligomer

A

S. cerevisiae **M S S P D D F E T A** P A E Y V D A L D P

K. lactis **M S H A D D F D T A** L D S P S V Q I K E

B

RAP1Nlys7 M K K K K K K K F S S P D D F E T A P A E Y V D A L D P

RAP1Δ15 M A L D P

RAP1Δ10 M A E Y V D A L D P

Figure 5-8 Rap1p N-Terminal Modifications **A**. Alignment of N-terminal sequences of *S. cerevisiae* and *K. lactis* Rap1p homologs. Seven of the first ten residues are identical. **B**. N-terminal sequences of three mutant discussed of Rap1p used in this work.

and monomer were observed, but not tetramers or dimers. This suggests that these oligomeric states of Rap1p are greatly destabilized by N-terminal modification.

We were very surprised to find that the N-terminally modified forms of Rap1p do not bind DNA at protein concentrations up to 1 μM. Even after the Rap1p N-terminal mutants have been incubated in 2 M urea and have the apparent molecular weight of monomers (as measured by size exclusion chromatography), they remain incapable of binding DNA below protein concentrations of 1 μM. However, at a concentration of 5 μM, urea-treated Rap1Δ10 shows evidence of DNA binding while untreated Rap1Δ10 shows none. Either only a small fraction of the urea-treated Rap1Δ10 is capable of binding DNA or the Rap1Δ10 monomer has much higher K_d than the Rap1p dimer. Perhaps in the monomeric state, the N-terminal domain is capable of masking the DNA binding domain.

To be sure that Rap1Δ10 has a functional DNA binding domain, we subjected Rap1Δ10 to proteolysis. We were surprised that Rap1Δ10 is quite resistant to proteolysis. Whereas Rap1p is completely proteolyzed to its three structural domains at a Rap1p:protease ratio of 800:1 at

4°C, Rap1Δ10 is completely resistant, even at a Rap1Δ10:protease ratio of 100:1 at 4°C. Proteolysis of Rap1Δ10 required that the protein be treated with 2 M urea, then proteolyzed at a Rap1Δ10:protease ratio of 500:1 at room temperature. This proteolyzed Rap1Δ10, when used in a gel shift experiment, yielded a shifted DNA band with a mobility identical to that of the Rap1 DNA binding domain. No DNA binding activity was recovered from proteolyzed Rap1Δ10 that had not been treated with urea. This shows that Rap1Δ10 has a fully functional DNA binding activity which is blocked in the aggregated and monomer forms of the protein. Rap1Δ10 DNA binding activity can be recovered by limited proteolysis of monomeric Rap1Δ10.

Perhaps when Rap1p is not competent to dimerize, it adopts a very stable, alternate conformation in which the DNA binding activity is masked by the N-terminal domain. Interaction *in trans* between the N-terminal domain of one molecule and the DNA binding domain of another might account for the observation of a larger aggregate.

Bacteriophage T4 Gene 32 protein provides a

precedent for this combination of findings. Gene 32 protein has been shown to contain a domain that masks its DNA binding activity. Oligomerization displaces this domain in a way that unmasks the DNA binding potential. As with Rap1p, the monomer form of the protein is highly resistant to proteolysis (Kowalczykowski et al., 1981; Casas-Finet and Karpel, 1993).

We have presented evidence that Rap1p is a dimer/tetramer *in vitro* based on its apparent molecular weight observed by gel filtration column analysis and Ferguson analysis of native gels. The oligomeric form of Rap1 is observed in full length Rap1 and structural domains containing the N-terminal domain (Rap1 NTD and Rap1 NTD+DBD). Rap1p structural domains that do not contain the N-terminal domain have the apparent molecular weight of a monomer (Rap1 DBD and Rap1 DBDTEL). From these findings we conclude that the Rap1p N-terminal domain functions as a required mediator of oligomerization. We found that deletion of as few as 9 residues from the N-terminus of Rap1 (RapΔ10) has dramatic effects on the protein *in vitro*. The modified protein appears to form an aggregate which can be

disrupted to a monomeric form by incubating the protein in 2 M urea. This observation of a Rap1p monomer makes an important contribution to our argument that Rap1p exists as an oligomer. In addition, RapΔ10 is not capable of binding its cognate DNA in either the aggregated or monomer form up to a concentration of 1μM. However, DNA binding activity is at least partially recovered at high RapΔ10 concentrations by urea treatment and by limited proteolysis of urea-treated RapΔ10. We suggest that when the N-terminal domain is present, dimerization is required for wild-type DNA binding. In other words, dimerization "unmasks" the Rap1p DNA binding activity *in vitro*.

Given this evidence that Rap1p functions as a dimer/tetramer, we must now consider a model that incorporates these new findings with what has previously been reported about Rap1p. Figure 5-9 shows a model in which Rap1p plays a primary role in assembling a higher order structure at the telomere. In this model, the binding of Rap1p to sequential sites in the telomere induces curvature of the DNA. The two subunits of a Rap1p dimer bind to sites which are brought into

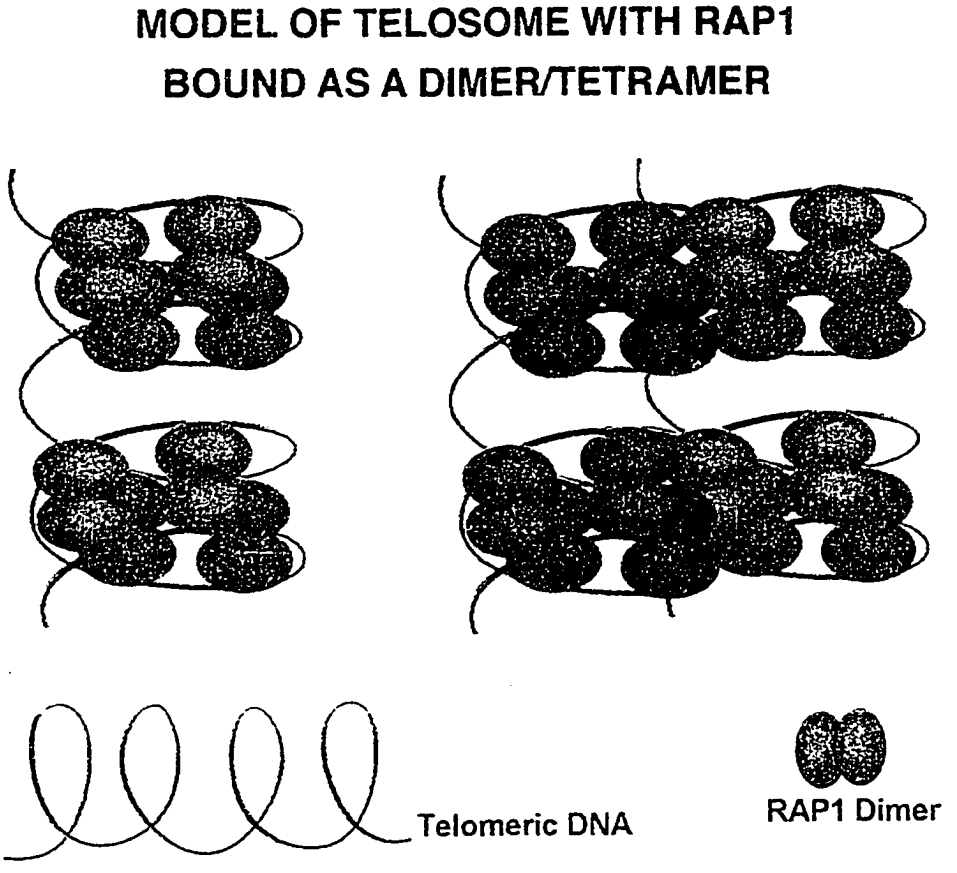


Figure 5-9 In this model, Rap1p binds as a dimer/tetramer to the double stranded region of yeast telomeres. The subunits of a Rap1p oligomer bind to DNA sites that are not adjacent, but are brought close in space by Rap1p DNA bending activity. Rap1p oligomerization may also play a role in mediating telomere-telomere interactions.

proximity by DNA bending but are not adjacent to one another. We suggest that Rap1p binds approximately every 25 base pairs, based on the work reported in Chapter III, and the binding of each subunit induces a >50° bend (Muller et al., 1994), thus initiating assembly of a higher order chromatin structure. Sir3p and Sir4p, which are required for maintaining the highly condensed structure at yeast telomeres, associate with the complex via Rap1p. The finding that Rap1p is an oligomer thus offers a molecular explanation for its role as a primary structural protein at yeast telomeres.

In diploid *S. cerevisiae* cells, telomeres are seen clustered at the nuclear periphery in many fewer foci than there are telomeres when localized by fluorescence *in situ* hybridization (Gotta et al., 1996), suggesting telomere-telomere association and organization. Including telomeric DNA sequence on a circular plasmid improves plasmid segregation in a Rap1p-dependent fashion (Longtine et al., 1993), suggesting that Rap1p binding sites in the plasmid provides a subnuclear localization that results in a more efficient segregation of the plasmid between mother and daughter cell. Rap1p

oligomerization may play an important role in mediating telomere-telomere association.

With the information that the N-terminal mediates oligomerization of Rap1p, we decided to now re-examine the effects of deleting the N-terminal domain as well as smaller N-terminal deletions on the functions of Rap1p *in vivo*.

Chapter VI

Deletion of Rap1p N-terminal Domain Affects Telomere Position Effect, Rap1p Immunolocalization, Mating Efficiency, and Telomere Length

Introduction

Rap1p is essential for viability in *Saccharomyces cerevisiae*. Rap1p activates transcription at a wide variety of genes, including many glycolytic enzyme and ribosomal protein genes, and *MAT*α1 and *MAT*α2 (Giesman et al., 1991; Huet et al., 1985). Rap1p also contributes to transcriptional silencing of the silent mating-type loci and to the telomere position effect (the transcriptional silencing of telomere-proximal genes) (Aparicio et al., 1991; Buchman et al., 1988a; Buchman et al., 1988b; Shore and Nasmyth, 1987; Shore et al., 1987). In addition, Rap1p is a telomere binding protein that plays a role in maintaining telomeres at a relatively constant length which is strain-specific (Berman et al., 1986; Lustig et al., 1990).

Mutational analysis has revealed that the central region (residues 361-596) is involved in DNA binding (Henry et al., 1990). Temperature sensitive alleles map primarily to the DNA binding region and phenotypes are almost certainly due to significantly reduced Rap1p DNA binding activity at the semipermissive temperature. These alleles display growth and mating defects. At semipermissive temperatures, decreases in *MAT*α steady-state transcript levels have been detected (Kurtz and Shore, 1991). These temperature sensitive alleles also have shortened telomeres at the semipermissive temperature (Lustig et al., 1990).

Mutational analysis has also revealed that the C-terminal region of Rap1p is required for transcriptional activation and silencing as well as telomere length regulation. Point mutations and deletions in the C-terminal region of Rap1p have a variety of phenotypes. Strains bearing various point mutations in the C-terminal region of Rap1p including R747S, G726E and D727N, D727A, and G726E display defective silencing at the silent mating-type loci (Kurtz and Shore, 1991; Shore and Nasmyth, 1987; Sussel and Shore, 1991). Strains from

which the C-terminal-most 144 or 165 residues of Rap1p have been deleted show a loss of telomere position effect (Kyrion et al., 1993) and have telomeres that are longer than normal (Kyrion et al., 1992).

Immunofluorescence microscopy of yeast spheroplasts probed with antibody to Rap1p reveals punctate perinuclear staining which colocalizes with telomeres in ~8 foci per diploid nucleus (Gotta and Gasser, 1996; Gotta et al., 1996; Palladino et al., 1993). This wild-type pattern is lost in a strain bearing *rap1-17* (Δ663-827) in which telomere position effect is lost and telomeres are drastically elongated (Cockell et al., 1995). However, a strain bearing *rap1-12* (G726E, D727N), which has increased telomere length and increased telomere position effect, the punctate perinuclear staining of a wild-type strain is seen. From this sampling, it appears that wild-type Rap1p localization is tied to wild-type telomeric silencing but not to telomere length regulation (Cockell et al., 1995). In fact, in *sir3Δ* and *sir4Δ* strains, telomeric silencing and wild-type Rap1p immunolocalization are lost, but telomeres are slightly shorter than in wild-type strains (Palladino et

al., 1993).

Because of our finding that the N-terminal domain of Rap1p mediates Rap1p dimerization and tetramerization *in vitro*, we were interested in what phenotypes, with respect to growth, mating efficiency, telomere position effect, telomere length and immunolocalization, we might find from deletion of the N-terminal domain of Rap1p. It has been reported that the N-terminal region of Rap1p (Δ43-279) is dispensable for viability and has no silencing phenotype (Moretti et al., 1994; Shore, 1994). However, we have observed the following phenotypes in strains bearing *rap1Δ10* or *rap1DBDTEL*: diminished telomeric silencing, enhanced mating by *MATα* cell types, slight telomere shortening and altered Rap1p immunolocalization.

Materials and Methods

Plasmid construction

pLP62/RAP1

The *Pst* I-*Xba* I fragment of pLP11 (nucleotides 858-3670) which contains the coding region for amino acids

V32-L827 as well as the *RAP1* terminator, was ligated into the *Pst* I-*Xba* I fragment of pLP62 to create pLP62/RAP 858-3670. A DNA fragment containing nucleotides 1-858 of the *RAP1* gene, which include the *RAP1* promoter and the coding region for amino acids M1-A31, with a *Xho* I restriction site inserted 5' of the first nucleotide was produced from pLP11 using PCR, with the following primers.

RNT CCT TCC CTT CTC GAG ATC TGG TAT TAC CGG T

RCT TCC TTC CTT CCT CTG CAG CAG AAC CAC TG

pLP62/RAP 858-3670 and this PCR product were digested with *Xho* I and *Pst* I and ligated together to create pLP62/RAP1 which includes the entire *RAP1* gene flanked by *Xho* I and *Xba* I restriction sites.

pLP62/RAP1Δ10

A DNA fragment containing the *RAP1* promoter plus codons for M1 and A12-A31 (i.e. deleting the codons for S2-P11) with a *Xho* I restriction site inserted 5' of the first nucleotide was produced by overlapping PCR mutagenesis. The 5' PCR product which includes nucleotides 1-767 (the promoter and codon for M1) and nucleotides 798-816 (coding for A12-A17) for overlap was

produced from pLP62 using the following primers.

RNT CCT TCC CTT CTC GAG ATC TGG TAT TAC CGG T

R3' ATG CAT CAA CAT ATT CTG CCA TAT TGA GAT AAT TCT GTAC

The 3' PCR product including nucleotides 798-858 (coding for A12-A31) and nucleotides 748-767 for overlap was produced from pLP11 using the following primers.

R5' GTA CAG ATT ATC TCA ATA TGG CAG AAT ATG TTG ATG CAT

RCT TCC TTC CTT CCT CTG CAG CAG AAC CAC TG

The PCR product incorporating the full *RAP1* promoter and coding for M1 and A12-A31 was produced from the mixture of the 5' and 3' PCR products using the following primers.

RNT CCT TCC CTT CTC GAG ATC TGG TAT TAC CGG T

RCT **TCC TTC CTT CCT CTG CAG CAG AAC CAC TG**

pLP62/RAP 858-3670 and this PCR product were digested with *Xho* I and *Pst* I and ligated together to create pLP62/RAPΔ10.

pLP62/RAP DBDTEL

The 8700 bp product of an *Nsi* I partial digest of pLP62/RAPΔ10 was religated to create pLP62/RAP1 DBDTEL. The correct product was selected for by digestion with *Nru* I. This construct includes the *RAP1* promoter

followed by the following sequence.

M	A	E	Y	V	D	A	S	F	Q	A	Q	R	S	M
ATG	GCA	GAA	TAT	GTT	GAT	GCA	TCT	TTT	CAA	GCA	CAA	AGG	TCC	ATG
1	12	13	14	15	16	17	342	343	344	345	346	347	348	349

Strain construction

RAP1 knockout

One copy of the *RAP1* open reading frame, from the ATG start codon through the codon for D825, was replaced by the *HIS3* gene using homologous recombination in two diploid yeast strains LPY 1569 and LPY 2553.

The template for the PCR reaction was a restriction fragment of pLP60 containing the *HIS3* gene. Primers (shown below) were as follows. RAP1-PROHIS contains 41 bases identical to *RAP1* 724-764 in the promotor region of the *RAP1* gene plus the 17 nucleotides from *HIS3* promoter, and RAP1-TERMHIS contains 54 bases complimentary to nucleotides 3238-3291, (including codons for D825-L827, the stop codon and five nucleotides of the terminator) plus the 17 nucleotides from the *HIS3* terminator (Baudin et al., 1993).

RAP1-PROHIS CGC CCT ACA TAA GAC ACT ATT TGC CTA CAG ATT ATC TCA ATC TCT TGG CCT CCT CTA G

RAP1-TERMHIS GGA GTA AAA TAA GTT AAA CAA TGA TGT TAC

TTA ATT CAA TTA CTC ATA ACA GGT CGT TCA GAA TGA CAG G

The PCR fragment (~1 kb) was gel purified and then transformed into LPY1569 and LPY2553 cells. Cells were made competent for transformation as follows. Two ml YPD (20 g/L peptone, 10 g/L yeast extract, 2% glucose) were inoculated with a single colony and cells grown overnight at 30•C. 50 ml YPD were inoculated from the overnight culture to an OD_{600} of 0.15 and cells were grown at 30•C to an OD_{600} of ~0.8. Cells were harvested by centrifugation, resuspended in 5 ml 1X TEL (100 mM lithium acetate, 4 mM Tris (pH 8.0), 1 mM EDTA) and shaken gently at room temperature overnight. Cells were harvested by centrifugation and resuspended in 0.5 ml 1X TEL. Transformation was performed as follows. 5 •l salmon sperm DNA were mixed with 10 µl PCR product, then 100 µl competent cells were added, mixed gently and incubated at room temperature for 30 minutes. 700 •l 40% PEG/1X TEL were added gently and the cells incubated at room temperature for 45 minutes. Cells were incubated at 42•C for 5 minutes and harvested by centrifugation (15-30 seconds at 5K). The supernatant was removed and cells resuspended in 150 µl 1 M sorbitol. The entire

transformation was spread onto his$^-$ media and incubated at 30°C. All transformants were struck for singles.

Proper placement of the *HIS3* gene was confirmed by PCR on genomic DNA from His$^-$ isolates using the following primers.

RNT CCT TCC TTC CTT CTC GAG ATC TGG TAT TAC CGG T

HIS-CHK ATG TGA TAA TGC CAA TCG CT

Genomic DNA was prepared as follows (Hoffman and Winston, 1987): 10 ml cells grown to saturation were harvested by centrifugation, resuspended in 0.5 ml water, and harvested by centrifugation for 5 seconds. Supernatant was poured off and cells were resuspended in residual liquid by vortexing 20-30 seconds. 200 •l Winston buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris (pH 8.0), 1 mM EDTA), 200 µl phenol-chloroform and 0.3 g acid washed beads were added. This was vortexed for 3 minutes, 200 µl TE added and the mixture centrifuged for 5 minutes at 13 K. The supernatant was transferred to a new tube and DNA precipitated by addition of 1 ml 100% ethanol. Precipitate was collected by centrifugation for 2 minutes, then resuspended in 400 µl TE, 30 µg Rnase A. This was incubated at 37°C for 5 minutes. DNA was

precipitated by addition of 10 µl 4 M ammonium acetate and 1 ml 100% ethanol and collected by centrifugation for 2 minutes. After drying, the pellet was resuspended in 20 µl TE. 1 µl of a 1:10 dilution was used for PCR.

The *RAP1* knockout was confirmed by inducing sporulation in transformants and dissecting tetrads. Since *RAP1* is an essential gene, dissection should yield 2 live spores and 2 dead. The 2 live spores should be his^-. Single colonies were patched to YPD and grown for 24 hours at 30°C. Sporulation medium (30.6 mM potassium acetate, 0.02% raffinose, 114 µM zinc acetate) was inoculated with these cells and incubated at 30°C for 3 days. Tetrads were dissected on YPD.

Diploid *RAP1* knockouts were transformed with pLP62, pLP62/RAP1, pLP62/RAP1Δ10 and pLP62/RAP1DBDTEL as described above. Transformations were spread on his^-, leu^- medium. From each transformation, 8 colonies were struck for singles. Single colonies were patched to YPD and sporulated by incubation in sporulation medium. Tetrads were dissected and genotypes of tetrad products were determined by replication to selective media. Cell lines determined to be HIS^+, LEU^+, and to mate and

segregate properly were suitable for further experiments.

Dilution assay

This assay was performed on cell lines derived from LPY1569 which contains a *URA3* marker at the telomere of the Left arm of chromosome VII. Cell were grown to saturation in YPD (no selection was necessary because the *HIS3* gene is integrated and the *LEU2* plasmid is required for viability). The OD_{600} of the overnight growth was determined and cells were diluted to OD_{600} of 1.0 in sterile water, then five fold serial dilutions were performed. Each set of dilutions was printed to a 5-FOA plate, a ura$^-$ plate, and a YPD plate, and incubated at 30°C. Duplicate control plates were incubated at 16°C and 37°C to evaluate temperature sensitivity.

Quantitative Mating Assay

Cells were grown to saturation, then diluted into 5 mL YPD to OD_{600} = 0.1 and grown to OD_{600}=0.5. These cells were diluted in YPD to 1:50, 1:2,000 and 1:10,000. 200 •l of the 1:10,000 dilution were plated onto YPD. 100 µl of undiluted cells, 1:50 dilution or 1:2,000 dilution were mixed with 100 µl of the appropriate mating tester and spread onto minimal medium (6.7 g/L yeast nitrogen

base without amino acids, 2% glucose). After incubation at 30°C for 3 days, colonies were counted. Mating efficiency was calculated as the number of colonies produced by mating divided by the number of colonies grown on YPD.

Assessment of *ADE2* transcription from telomere-proximal locus.

YPD plates from the quantitative mating assay were stored at 4°C for one week. Red colonies, white colonies and sectored colonies were counted for each plate.

Immunofluorescence staining

A saturated culture was diluted into 50 ml YPD and grown overnight at 30°C to OD_{600} of 1.0. Cells were harvested by centrifugation at 1200 x G for 5 minutes and resuspended in 1 ml 0.1 M EDTA/KOH, 10 mM DTT per 0.1 g cells. These cells were incubated at 30°C for 10 minutes with gentle agitation, then collected by centrifugation at 800xg for 5 minutes. Cell were resuspended in 1 ml YPD/1.2 M sorbitol per 0.1 g cells. Lyticase was added to a final concentration of 1000U/ml and zymolyase to a final concentration of 0.4 mg/ml. Cells were incubated at 30°C with gentle agitation. The progress of

spheroplast formation was monitored by examining the cells under the microscope every 5 minutes. Spheroplasts were fixed as follows. Paraformaldehye was added to spheroplasts to a final concentration of 4% and spheroplasts were incubated at room temperature for 20 minutes with occasional gentle inversion. Five volumes of YPD/1.2 M sorbitol were added and spheroplasts harvested by centrifugation at 800xg. Spheroplasts were washed twice in 20 ml YPD/1.2 M sorbitol. Spheroplasts were resuspended in 3 ml YPD per 0.4 g original cell pellet, then introduced to an 8-well teflon slide. After 2 minutes, liquid was removed from each well and the slide air dried for 2 minutes. The slide was placed in a -20•C methanol bath for 6 minutes, in a -20°C acetone bath for 1 minute, then air dried for 3 minutes. The slide was incubated in PBS/0.1% Triton X-100/1% ovalbumin at room temperature for 20 minutes, with gentle shaking 2-3 times during the incubation. Each well was covered with 5 µl of affinity purified antibody to Rap1p (Enomoto et al., 1997) in PBS/0.1% Triton X-100/1% ovalbumin, the slide was sealed with a cover slip and incubated in a humid chamber at 4°C overnight. The slide was washed 5

times in PBS/0.1% Triton X-100 at room temperature. Secondary antibody was preadsorbed on yeast spheroplasts as follows. Spheroplasts were harvested and washed 3 times in 1 ml PBS, then resuspended in 200 µl PBS. Two µl FITC-conjugated goat anti-rabbit or 2 µl Texas red-conjugated goat anti-rabbit were added to 100 µl washed spheroplasts, wrapped in foil and incubated on a rotating shaker at 4°C for 1 hour. Spheroplasts were pelleted, supernatant removed to a clean tube and Triton X-100 added to a final concentration of 0.1%. Ten µl preadsorbed secondary antibody were added to each well and incubated at 37°C in the dark for 1 hour. Keeping the slide in the dark as much as possible, it was washed twice in PBS/0.1% Triton X-100, then 2 µl DAPI (diluted 1:1000 in PBS/0.1% Triton X-100) were applied to each well and incubated for 3 minutes. The slide was washed twice more in PBS/0.1% Triton X-100, then dried carefully, mounted in antifade solution (24 mg/ml DABCO in 87% glycerol, pH7.5) and sealed with a coverslip. The stained spheroplasts were visualized on a Zeiss III RS fluorescence microscope and images were captured using MetaMorph Software. Results were quantified by counting

foci visible in each spheroplast image captured and calculating an independent means t-test comparing each mutant to wild-type.

Determination of telomere length

This experiment was performed by Joanna Lowell in Lorraine Pillus' lab as follows. Genomic DNA was prepared exactly as described above (Hoffman and Winston, 1987) and resuspended in 100 µl TE. 25 µl of genomic DNA were digested with *Xho* 1 restriction endonuclease and run on a 0.7% agarose TAE (40 mM Tris-acetate, 1 mM EDTA) gel. Transfer of DNA to Hybond N+ was accomplished as follows. The gel was incubated twice for 10 minutes in 0.25 M HCl, rinsed once briefly in water and then for 5 minutes in water. The gel was then gently shaken in 0.4 M NaOH for 20 minutes. The DNA was then transferred to Hybond N+ by wicking for 2 hours (Ausubel et al., 1998). For hybridization, the blot was prewet in 5X SSPE (.01875 N NaOH, 75 mM NaCl, 50 mM NaH_2PO_4, 5 mM EDTA). Blot was then incubated with 10 ml prehybridization mix (4X SSPE, 200 µg/ml Ficoll, 200 µg/ ml polyvinylpyrrolidone, 200 µg/ ml BSA, 0.1% SDS, 100 µg/ml salmon sperm DNA) for 30 minutes at 65°C.

Labeled probe was prepared as follows. Plasmid pLP 127 (published as pYLPV) (Gilson et al., 1993) as digested with *Eco*R I restriction endonuclease and 350 bp fragment was isolated and labeled with α^{32}P-dCTP using the Amersham Rediprime kit. Incorporated label was separated from unincorporated using a G-50 spin column poured in a 1 ml syringe. Labeled probe was added to prehybridization mix and allowed to hybridize at 65°C overnight. Probe was removed and blot washed in 2X SSPE and 0.1% SDS 1 time at room temperature. It was washed with 0.2X SSPE with several changes over 2 hours at 65°C. Blot was then exposed.

Results

RAP1 knockout

One copy of the *RAP1* gene was replaced with *HIS3* in two diploid yeast strains: LPY 1569 and LPY 2553. This was confirmed by PCR on genomic DNA from His^+ colonies recovered from transformation. The 5' primer was complimentary to a region of the *RAP1* promoter and the 3' primer was complimentary to a region of the *HIS3* gene.

8/14 LPY 1569 transformants yielded the appropriate PCR product. Two of these isolates were induced to sporulate and tetrad dissection yielded 2 live: 2 dead as expected when an essential gene is knocked out. 1/1 LPY 2553 transformants yielded the appropriate PCR product. This isolate was induced to sporulate and tetrad dissection yielded 2 live: 2 dead as expected. From both cell lines, live cells were his^-.

***rap1Δ10* and *rap1DBDTEL* viability, temperature sensitivity**

Haploid strains in which rap1Δ10p or rap1DBDTELp is the only form of Rap1p present are viable. They have no apparent growth defects and are not sensitive to heat or cold. They grow as wild-type at 16°, 30°C, and 37°C.

Telomere position effect

Telomere position effect on expression of the *URA3* gene placed at the telomere on chromosome VIIL was assessed by resistance to 5-FOA which kills cells that are transcribing *URA3*. The *URA3* gene product metabolizes 5-fluor-orotic acid to 5-fluoro-uracil monophosphate, which is lethal to yeast cells (Boeke et al., 1984). Strains expressing rap1Δ10p or rap1DBDTELp are more sensitive to 5-FOA than wild-type cells, indicating

increased transcription of the telomere-proximal *URA3* gene (see Figure 6-1). Quantitative assessment reveals that strains expressing rap1Δ10p are 70 times more sensitive to 5-FOA than wild-type cells. Strains expressing rap1DBDTELp are 28 times more sensitive to 5-FOA than wild-type cells (see Table 6.1). These results indicate that in *rap1Δ10* and *rap1dbdtel* strains, telomeric silencing is diminished as measured by transcription of a telomere-proximal *URA3* gene.

The telomere at chromosome VR is marked with *ADE2* in strains derived from LPY1569. Transcription of *ADE2* can be assessed by evaluating colony color. In an *ade2* mutant or when *ADE2* is transcriptionally silenced, a pigmented biosynthetic precurser of adenine accumulates within cells, resulting in red colonies. When the *ADE2* gene is expressed, colonies are white. When the transcriptional state of *ADE2* is switching between silenced and expressed, red and white sectors within one colony are observed. The fraction of white colonies from cells expressing wild-type Rap1p, Rap1Δ10p and Rap1DBDTELp is reported in Table 6.2. The remainder of the cells counted contained red and white sectors; none

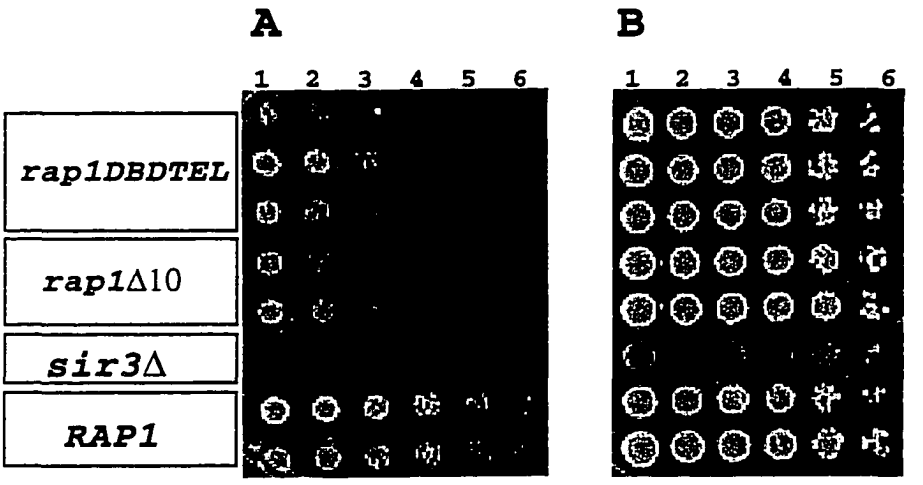


Figure 6-1 Growth of *rap1DBDTEL*, *rap1Δ10*, *sir3Δ*, and *RAP1* strains on (A) 5-FOA and (B) YPD (**A**) *rap1DBDTEL* and *rap1Δ10* strains are considerably more sensitive to 5-FOA than *RAP1* strains, indicating that there is partial derepression of a telomere-proximal *URA3* gene in these mutant strains. A *sir3Δ* strain is included as a representation of complete derepression of a telomere proximal URA3 gene. Shown are five-fold dilutions from OD_{600} =1.0 in lane 1. (**B**) Growth on YPD is equal for all strains.

Table 6.1 Growth on 5-FOA This table lists the number of dilutions to which the strain grew on 5-FOA (mean fold dilution), the standard deviation of the mean and the statistical significance of the differences. The differences between growth of *RAP1* and *rap1DBDTEL* and *RAP1* and *rap1Δ10* on 5-FOA are highly significant. The difference between the sensitivity to 5-FOA of *rap1DBDTEL* and *rap1Δ10* strains is also significant.

strain	# of cases	mean fold dilution (5^x)	standard deviation	statistical significance
RAP1	9	5.4	.53	$\|t\|=7.1$ $p<.001$
rap1DBDTEL	29	3.4	.82	
RAP1	9	5.4	.53	$\|t\|=9.8$ $p<.001$
rap1Δ10	10	2.8	.63	
rap1DBDTEL	29	3.4	.82	$\|t\|=2.0$ $p=.050$
rap1Δ10	10	2.8	.63	

Table 6.2 Expression of a Telomere-Proximal *ADE2* Reporter Gene Reported here is the mean fraction of colonies in which transcription of the *ADE2* gene at telomere VR is completely derepressed, the standard deviation of the mean (S D) and the statistical significance of these differences. The differences in the fraction of cells completely derepressed at telomere VR between *RAP1* and *rap1DBDTEL* and between *RAP1* and *rap1Δ10* are highly significant. The difference between the mean fraction of colonies in which transcription of the *ADE2* gene at telomere VR is completely derepressed in strains expressing *rap1DBDTEL* and *rap1Δ10* is also highly significant.

strain	# of isolates	colonies evaluated	Fraction of colonies per isolate completely derepressed at *ADE2*	Standard Deviation	statistical significance
RAP1	6	1313	.25	.11	$\|t\|=5.3$ $p<.001$
rap1DBDTEL	13	2865	.68	.25	
RAP1	6	1313	.25	.11	$\|t\|=10.0$ $p<.001$
rap1Δ10	4	726	.95	.11	
rap1DBDTEL	13	2865	.68	.25	$\|t\|=3.00$ $p=.011$
rap1Δ10	4	726	.95	.11	

were completely red. 94.5% of colonies from cells expressing rap1Δ10p are white, 68% of colonies from cells expressing rap1DBDTELp are white, and 25% of colonies from cells expressing wild-type Rap1p are white. This indicates that both mutant forms of Rap1p diminish telomere position effect at telomere VR, with the deletion of only 10 amino acids having a more severe effect than deletion of the entire N-terminal domain.

Western Blot

A western blot of whole cell extracts from strains expressing wild-type Rap1p and rap1Δ10p, probed with antibody to Rap1p, showed that rap1Δ10p is expressed at levels equal to wild-type Rap1p.

Immunofluorescence

RAP1, *rap1dbdtel*, and *rap1Δ10* cells probed with antibody to Rap1p and stained with fluorescently-tagged secondary antibody are shown in Figure 6-2. Qualitatively, it appears that wild-type cells have the pattern of punctate staining described in the literature (Cockell et al., 1995; Gotta et al., 1996; Palladino et al., 1993) for *RAP1* strains, *rap1Δ10* cells have a

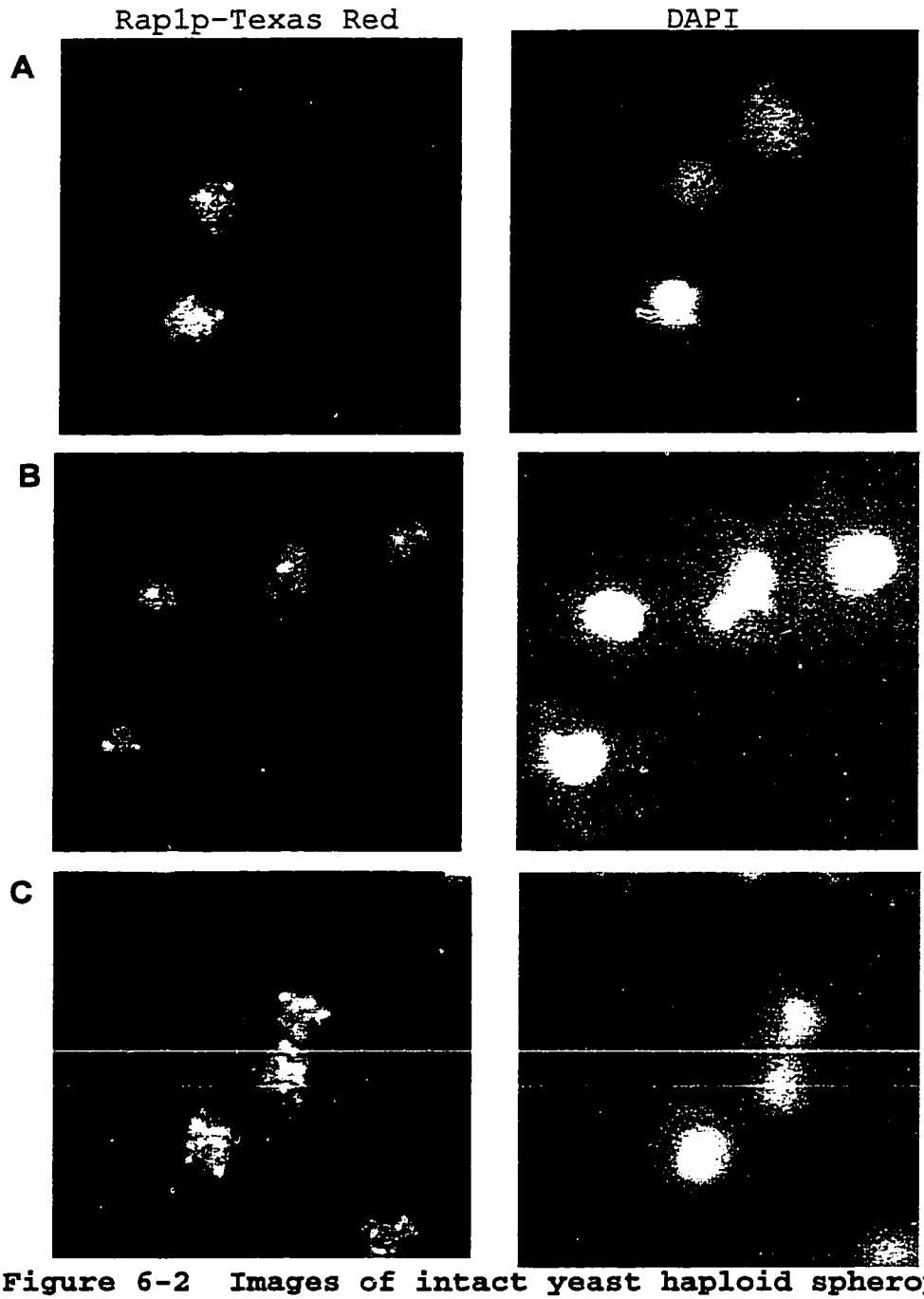


Figure 6-2 Images of intact yeast haploid spheroplasts probed with antibody to Rap1p and stained with a rhodamine-tagged secondary antibody and counterstained with DAPI to identify the nucleus. (A) A *RAP1* strain gives a pattern of punctate staining with a mean of 3.5 foci per spheroplast. (B) A *rap1DBDTEL* strain gives a pattern of punctate staining, but with fewer foci (a mean of 2.6 foci per spheroplast). (C) A *rap1*Δ10 gives a pattern of punctate staining similar to wild-type, but with more foci (a mean of 4.2 foci per spheroplast).

staining pattern similar to wild-type and *rap1dbdtel* cells have fewer foci, some of which are larger than wild-type. When the number of foci were counted for ~30 cells each of wild-type and each mutant, there were fewer foci in cells expressing rap1DBDTELp and more foci in cells expressing rap1Δ10p than in wild-type cells. These results are reported in Table 6.3.

Telomere length

A Southern blot of telomeric DNA is shown in Figure 6-3. It appears that the telomeres of cells expressing rap1DBDTELp are slightly shorter than wild-type. The telomeres of cells expressing rap1Δ10p appear to be wild-type length.

Mating

No mating defects were observed in cells expressing rap1Δ10p or rap1DBDTELp. Results of a t-test to evaluate the statistical significance of any differences in mating efficiency among cells expressing wild-type Rap1p, rap1Δ10p, and rap1DBDTELp are reported in Table 6.4. *MATα rap1DBDTEL* strains, mated ~60% better than *MATa rap1DBDTEL*, *MATα RAP1*, or *MATa RAP1* strains. The *p* values for these

Table 6.3 Number of Rap1p Foci per Cell Visible Using Immunofluorescence Microscopy As measured using immunofluorescence microscopy, significantly fewer Rap1p foci were visible in a strain expressing rap1DBDTELp than in a wild-type strain. More Rap1p foci per cell were visible in strains expressing rap1Δ10 than in wild-type. The *p* value for this difference of .061 just misses the traditional cutoff for statistical significance of 0.05.

strain	number of cells	mean number of foci	standard deviation	statistical significance
RAP1	31	3.5	1.3	$\|t\|=2.6$ $p=.012$
rap1DBDTEL	29	2.6	1.5	
RAP1	31	3.5	1.3	$\|t\|=1.9$ $p=.061$
rap1Δ10	33	4.2	1.4	

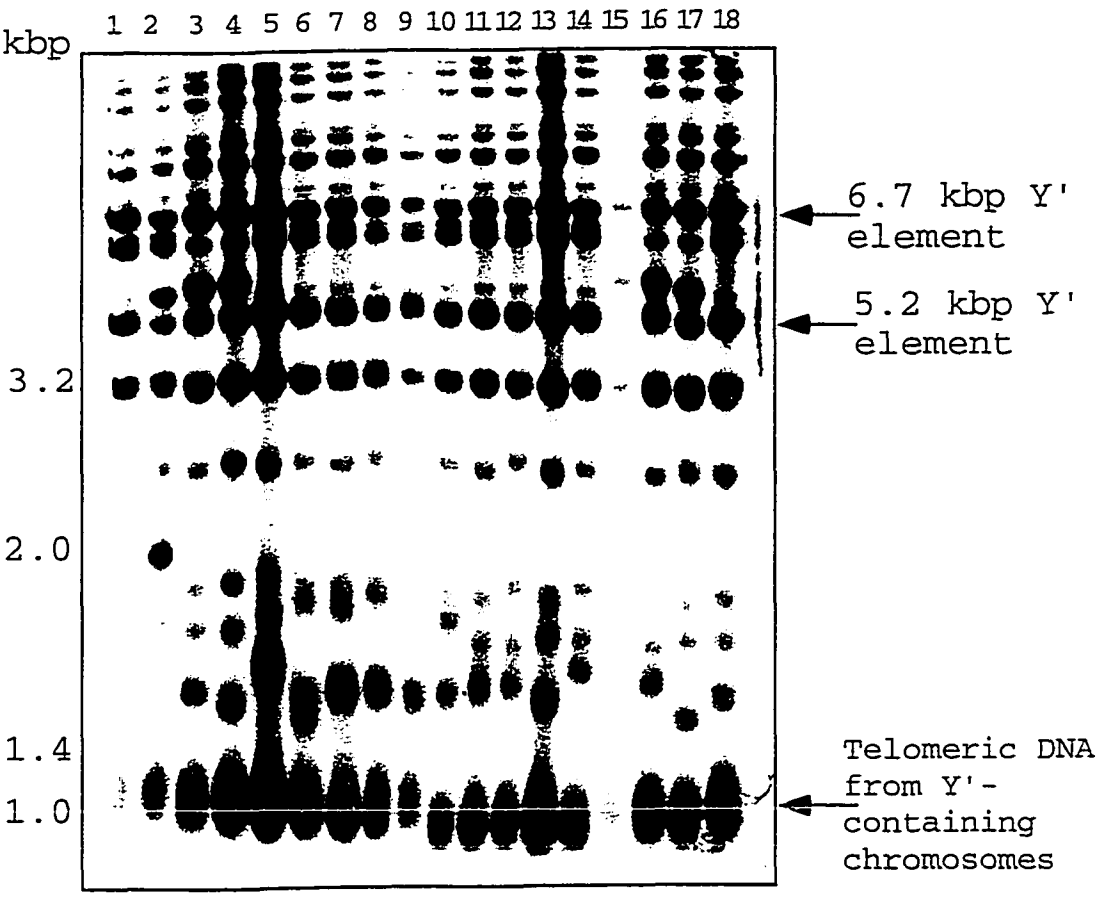


Figure 6-3. Southern blot of genomic DNA cleaved with *Xho* I restriction endonuclease and probed with the CA strand of the yeast telomeric repeat Lanes 1 and 18, WT; lane 2, *sir3*Δ; lanes 3, 4, 15-17, *rap1*Δ covered by *RAP1* on a plasmid; lanes 5-9, *Rap1*Δ*10* ; lanes 10-14, *rap1DBDTEL*. The telomeres appear to be slightly shorter in isolates of a *rap1DBDTEL* strains.

Table 6.4 Mating Efficiency This table reports only differences in mating efficiency that meet (or nearly meet) the traditional test of statistical significance (p=.05). α cell types expressing Rap1p DBDTEL mate at ~60% higher efficiency than wild-type cells or **a** cell types expressing Rap1 DBDTEL.

cell type strain	number of cases	mean mating efficiency	standard deviation	statistical significance
*MAT*α DBDTEL	8	.37	.16	|t|=2.4 p=.04
*MAT***a** DBDTEL	8	.23	.05	
α DBDTEL	8	.37	.16	|t|=2.9 p=.02
a wild-type	3	.20	.04	
α DBDTEL	8	.37	.16	|t|=2.1 p=.07
α wild-type	3	.23	.07	

differences are at or very near the traditional cutoff for statistical significance of 0.05.

Discussion

Rap1p is a *trans*-activator of many growth-related genes (Huet et al., 1985) and of the Matα genes, which specify the α cell type (Giesman et al., 1991; Kurtz and Shore, 1991). Rap1p binds to the silencer elements of the silent mating-type loci, which must be maintained in a transcriptionally repressed state for the cell to mate normally (Shore and Nasmyth, 1987). Rap1p is primarily localized to the telomeres as judged by immunofluorescence microscopy (Cockell et al., 1995; Gotta et al., 1996; Palladino et al., 1993; Shore and Nasmyth, 1987). Rap1p has two reported functions at telomeres. It plays a role in the transcriptional silencing of telomere-proximal genes known as telomere position effect (Gottschling et al., 1990) and it is involved in telomere length regulation (Lustig et al., 1990).

It is not surprising, given its many functions, that

mutations in Rap1p have been reported to have a wide variety of phenotypes. Temperature-sensitive alleles have mutations in the DNA binding domain (Kurtz and Shore, 1991). These have been reported to have severe growth phenotypes and shortened telomeres *in vivo* and diminished Rap1p DNA binding activity *in vitro* when grown at the semipermissive temperature. Mutations in the C-terminal region ofRap1p have been reported to affect silencing at the silent mating-type loci (Sussel and Shore, 1991), telomere length (Lustig et al., 1990), telomere position effect (Liu et al., 1994), and immunolocalization (Cockell et al., 1995). Large deletions from the N-terminal domain of Rap1p have been reported to have no phenotype (Moretti et al., 1994; Shore, 1994).

We found that Rap1p forms stable dimers and tetramers mediated by the N-terminal domain of the protein in solution and suggested a model in which these dimers and tetramers are important for assembling a higher order structure at yeast telomeres as well as in telomere-telomere associations. To test this model, we have constructed yeast strains containing a complete

deletion of the RAP1 gene which is covered by plasmid-borne *RAP1*, *rap1Δ10*, or *rap1DBDTEL*. We wished to ask if deletion of 10 residues from the N-terminus of Rap1p or deletion of the entire N-terminus would have any phenotypes with respect to growth, silencing at the silent mating-type loci, telomere position effect, telomere length or Rap1p immunolocalization. We have found that, in fact, expressing *rap1Δ10* or *rap1DBDTEL* in a *rap1Δ* strain affects telomere position effect, mating efficiency, telomere length and immunolocalization.

Strains expressing rap1Δ10p or rap1DBDTELp as the only source of Rap1p have no growth phenotype as judged by growth-rate and the fact that they show no temperature sensitivity. These cells grow as wild-type at 16•C, 30•C and 37•C. This agrees with published reports that the N-terminal domain of Rap1p is dispensable for viability (Shore, 1994).

To determine the effect of these *rap1* mutations on telomere position effect, we constructed strains derived from LPY1569, which has two marked telomeres. *URA3* has been placed adjacent to the telomere at chromosome VIIL and *ADE2* has been placed adjacent to the telomere at VR.

Since cells transcribing *URA3* will not grow on 5-FOA, growth on media containing 5-FOA provides a sensitive assay for the transcriptional state of the telomere-proximal *URA3* gene. We found that cells expressing rap1DBDTELp are 28 times more sensitive to 5-FOA than wild-type, while cells expressing rap1Δ10p are 70 times more sensitive to 5-FOA than wild-type. This effect on *URA3* expression depends on a wild-type copy of *PPR1*, a required *trans*-activator of *URA3* transcription (data not shown). Aparicio and Gottschling found that overcoming the silencing of a telomeric *URA3* gene required *PPR1* and passage through the cell cycle. They concluded that during replication, Ppr1p competes with the components of telomeric chromatin for access to the DNA (Aparicio and Gottschling, 1994).

In light of this, our results suggest that our N-terminal *rap1* mutants may be less successful than wild-type at assembling telomeric chromatin structure than wild-type. This could be due to their apparent inability to oligomerize, which could slow assembly considerably. These N-terminal *rap1* mutations could also result in a less highly condensed chromatin structure which might

improve the ability of Ppr1p to compete for access to the *URA3* promoter.

In our biochemical experiments, we found that rap1DBDTELp exists as a monomer but is capable of binding DNA comparably to wild-type Rap1p. In contrast, the *rap1Δ10* mutation lead to a large aggregate that did not bind DNA but which could be disrupted to yield monomeric rap1Δ10p. The more severe phenotype of *rap1Δ10* could result from a reduced DNA binding activity.

It is also possible that the reduction in Telomere Position Effect we have observed in *rap1Δ10* and *rap1DBDTEL* mutants may reflect impaired Rap1p interactions with other components of the telosome. Such protein-protein interactions involving the N-terminal domain of Rap1p have not been reported. However, when the 2-hybrid screen has been used to detect interactions between Rap1p and other yeast proteins, the N-terminal region of Rap1p has been deleted. A candidate for an interaction with the N-terminal domain of Rap1p is Rap1p Localization Factor 2 (*RLF2*) (Enomoto et al., 1997). *RLF2* is a subunit of yeast chromatin assembly factor-I, which does not localize to telomeres but is required for

the function and organization of telomeric chromatin *in vivo* (Enomoto et al., 1997).

The reduction in telomere position effect caused by mutations in the N-terminal domain of Rap1p is less severe those seen in a *sir3Δ* strain (Figure 6-1), where no 5-FOA-resistant colonies are seen.

Previously, it was reported that the deletion of amino acids 43-279 from Rap1p has no affect on transcription of *URA3* in a strain containing *URA3* at telomere VIIL (Moretti et al., 1994). Residues 1-43 may encompass the dimerization domain, in which case this result is consistent with our model that oligomerization is necessary for full Rap1p function. This could be tested by assessing the behavior of the *rap1Δ43-279* construct *in vitro* as well as *in vivo*. Alternatively, differences in strains used in the two studies could account for discrepancy in the result. The experiments using the *rap1Δ43-279* construct were performed in a strain where *LEU2* replaced the region of the *RAP1* gene coding for amino acids 1-760. The experiments using the *rap1Δ10* and *rap1DBDTEL* constructs were performed in a strain in which the region of the *RAP1* gene coding for

amino acids 1-824 has been replaced by the *HIS3* gene. In addition, expression of *rap1Δ43-279* was from a *HIS3* CEN plasmid, while expression of *rap1Δ10* and *rap1DBDTEL* was from a *LEU2* CEN plasmid.

We also assessed telomere position effect at chromosome VR, which contains a telomere-proximal *ADE2* gene, by counting red, white, and sectored colonies. Red colonies indicate that *ADE2* is transcriptionally repressed, white colonies indicate that *ADE2* is transcriptionally active, and sectored colonies indicate clonal switching in transcription state (*ADE2* transcription is either "on" or "off"). We again observed a reduction in telomere position effect in strains expressing rap1Δ10p and rap1DBDTELp. 25% of wild-type colonies are white, while 68% of colonies from cells expressing rap1DBDTELp are white, and 95% of colonies from cells expressing rap1Δ10p are white. The remainder of the colonies counted showed red and white sectors, indicating clonal changes in the transcriptional state of the telomere proximal *ADE2* gene at chromosome VR. This result indicates that these N-terminal mutations in Rap1p have similar effects on the structure

of telomere VIIL and VR.

Mating efficiency of *MAT***a** strains bearing either *rap1DBDTEL* or *rap1Δ10* was equal to wild-type, so it appears that these mutations in the N-terminal domain of Rap1p do not result in derepression of the silent mating-type loci. This should be further evaluated, however, in a strain bearing a partially disabled silencer. It has been shown that the *HM* silencers have redundancy, so that they function after deletion of the Rap1p binding site, the ABF1 binding site or the ARS consensus sequence. Several Rap1p mutations that affect *HM* silencing have been detected only in a partially disabled silencer background (Sussel and Shore, 1991).

Interestingly, mating efficiency is 60% better in *MATα rap1DBDTEL* strains than in *MAT***a** *rap1DBDTEL*, *MATα RAP1*, or *MAT***a** *RAP1* strains (see Table 6.4). *MATα rap1Δ10* strains mate at the same efficiency as *MATα RAP1* cells. We suggest that the increase in mating efficiency in *MATα rap1DBDTEL* may be due to a redistribution of Rap1p from the telomeres as a result of the inability of rap1DBDTELp to oligomerize. Since Rap1p is a *trans*-activator of *Matα1* and *Matα2* transcription, a reduction in the amount

of Rap1p bound to telomeres could shift more Rap1p to other of its binding sites. This hypothesis could be tested by probing for the level of *Matα1* and *Matα2* transcripts in this mutant compared to wild-type. A shift in Rap1p partitioning between two loci has been proposed by Buck and Shore. They suggest that in *rap1^s* mutants, which have longer than wild-type telomeres, silencing defects are the result of a shift in the balance of bound Rap1p away from silencers and toward telomeres (Buck and Shore, 1995).

Using immunofluorescence microscopy to localize Rap1p in haploid yeast spheroplasts, we observed the normal pattern of punctate staining described in the literature for wild-type cells (Cockell et al., 1995; Gotta et al., 1996; Palladino et al., 1993) (see Figure 6-2). In cells expressing rap1DBDTELp, however, there seem to be fewer foci than in wild-type cells and some of them are quite large. Cells expressing rap1Δ10p appear to have a pattern of staining similar to wild-type. To quantify these results, we counted the number of foci in ~30 cells each of *RAP1*, *rap1DBDTEL*, and *rap1Δ10* strains. It has been reported that a diploid yeast nucleus has

approximately 8 Rap1p foci (Gotta et al., 1996). In the haploid nuclei studied in these experiments, the mean number of foci in wild-type cells is 3.5, while in cells expressing rap1DBDTELp, the mean number of foci is 2.6. These differences are highly significant ($|t|=2.59$, $p=.012$). The mean number of foci in cells expressing rap1Δ10 is 4.2 as compared to the wild-type mean of 3.6 foci. The *p* value for this difference just misses the traditional cutoff for statistical significance of 0.05 ($|t|=1.91$, $p=.061$). Why are there fewer Rap1p foci in *rap1DBDTEL* strains than in wild-type and more foci in *rap1Δ10* cells than in wild-type? It is hard to interpret these data without additional information. In particular, has there been any change in telomere localization? This can be ascertained using fluorescent *in situ* hybridization of yeast telomeres. It would also be useful to repeat the experiment using diploid cells since their nuclei are larger and therefore easier to see in the microscope. It is possible that the appearance of fewer Rap1p foci in *rap1DBDTEL* cells reflects an increased background signal due to an increase in Rap1p molecules that are not localized to the telomeres.

Alternatively, there may be fewer Rap1p foci due to a relocalization of Rap1p. The appearance of more Rap1p foci in *rap1Δ10* may reflect a disruption of normal telomere-telomere interaction mediated by Rap1p. These findings are tantalizing in that they support our hypothesis that the N-terminal domain of Rap1p is important for the higher order structure of yeast telomeres, but more research is required to fully understand what they mean.

Joanna Lowell, a graduate student in the Pillus lab, investigated the telomere length of *rap1DBDTEL* and *rap1Δ10* strains (Figure 6-3). It appears that the telomeres of five isolates of a *rap1DBDTEL* strain are slightly shorter than wild-type. The telomeres of cells expressing rap1Δ10p appear to be the same length as wild-type. *sir3Δ* and *sir4Δ strains* also have shorter than wild-type telomeres (Palladino et al., 1993). The mechanism by which this change occurs has not been determined. Further experiments to clarify the telomere length phenotype of *rap1DBDTEL* and *rap1Δ10* strains are underway.

Strains expressing rap1Δ10p and rap1DBDTELp have

some things in common. Both display diminished telomeric silencing, but *rap1•10* results in more severe derepression than *rap1DBDTEL* at telomeres marked with a *URA3* gene and an *ADE2* gene. Immunofluorescence microscopy suggests that there are more Rap1p foci in *rap1•10* cells than in wild-type and fewer Rap1p foci in *rap1DBDTEL* cells than in wild-type. *MATα rap1DBDTEL* strains appear to mate at a higher efficiency than wild-type while *rap1Δ10* strains mate as wild-type. Finally, it seems that strains expressing rap1DBDTELp have slightly shortened telomeres.

Clearly, the N-terminal domain of Rap1p has a function *in vivo*. The results of our experiments are consistent with the idea that the N-terminal domain of Rap1p mediates oligomerization and that this oligomerization is important for assembly of a higher order telomeric chromatin structure. It is also possible that oligomerization of Rap1p may be required for interaction of Rap1p with other proteins that participate in the assembly of a higher order chromatin structure at yeast telomeres. It may also be that in addition to, or instead of, mediating oligomerization *in vivo*, the

extreme N-terminus of Rap1p interacts specifically with other proteins in assembling and maintaining telomeric chromatin structure.

Bibliography

Aparicio, O. M., Billington, B. L., and Gottschling, D. E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. Cell *66*, 1279-1287.

Aparicio, O. M., and Gottschling, D. E. (1994). Overcoming telomeric silencing: a *trans*-activator competes to establish gene expression in a cell cycle-dependent way. Genes Dev *8*, 1133-1146.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., D., Seidman, J. G., Smith, J., and Struhl, K. (1998). Current Protocols in Molecular Biology, V. Chanda, ed.: John Wiley and Sons).

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Research *21*, 3329-3330.

Berman, J., Tachibana, C. Y., and Tye, B.-K. (1986). Identification of a telomere-binding activity from yeast. Proc Natl Acad Sci U S A *83*, 3713-3717.

Boeke, J. d., LaCroute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotdine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet *197*, 345-346.

Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. A. (1985). Characterization of a "Silencer" in Yeast: A DNA Sequence with Properties Opposite to Those of a Transcriptional Enhancer. Cell *41*, 41-48.

Brand, A. H., Micklem, G., and Nasmyth, K. (1987). A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. Cell *51*, 709-719.

Brenowitz, M., Mandal, N., Pickar, A., Jamison, E., and Adhya, S. (1991a). DNA-binding Properties of a Lac Repressor Mutant Incapable of Forming Tetramers. J Biol Chem *266*, 1281-1288.

Brenowitz, M., Pickar, A., Jamison, E. (1991b). Stability of a *Lac* Repressor Mediated "Looped Complex" Biochemistry *30*, 5986-5998.

Buchman, A. R., Kimmerly, W. J., Rine, J., and Kornberg, R. D. (1988a). Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccharomyces cerevisiae*. Mol Cell Biol *8*, 210-225.

Buchman, A. R., Lue, N. F., and Kornberg, R. D. (1988b). Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. Mol Cell Biol *8*, 5086-5099.

Buck, S. W., and Shore, D. (1995). Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast. Genes Dev *9*, 370-84.

Carlson, N. G., and Little, J. W. (1993). Highly Cooperative DNA Binding by the Coliphage HK022 Repressor. J Mol Biol *230*, 1108-1130.

Casas-Finet, J. R., and Karpel, R. L. (1993). Bacteriophage T4 Gene 32: Modulation of Protein-Nucleic Acid and Protein-Protein Association by Structural Domains. Biochemistry *32*, 9735-44.

Cockell, M., Palladino, F., Laroche, T., Kyrion, G., Liu, C., Lustig, A. J., and Gasser, S. M. (1995). The carboxy termini of Sir4 and Rap1 affect Sir3 localization: evidence for a multicomponent complex required for yeast telomeric silencing. J Cell Biol *129*, 909-24.

Conrad, M. N., Wright, J. H., Wolf, A. J., and Zakian, V. A. (1990). RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and

decreases chromosome stability. Cell *63*, 739-50.

Drazinic, C. M., Smerage, J. B., Lopez, M. C., and Baker, H. V. (1996). Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). Mol Cell Biol *16*, 3187-96.

Efron, B. (1992). The jacknife, the bootstrap,and other resampling plans (Philadelphia: S. I. A. M.).

Enomoto, S., McCune-Zierath, P., Gerami-Nejad, M., Sanders, M. A., and Berman, J. (1997). RLF2, a subunit of yeast chromatin assembly factor-I, is required for telomeric function *in vivo*. Genes and Development *11*, 358-370.

Ferguson, K. A. (1964). Starch-Gel Electrophoresis-Application to the Classification of Pituitary Proteins and Peptides. Metabolism *13*, 985-1002.

Freeman, K., Gwadz, M., and Shore, D. (1995). Molecular and genetic analysis of the toxic effect of RAP1 overexpression in yeast. Genetics *141*, 1253-62.

Giesman, D., Best, L., and Tatchell, K. (1991). The role of RAP1 in the regulation of the MAT alpha locus. Mol Cell Biol *11*, 1069-79.

Gilson, E., and Gasser, S. M. (1995). Repressor Activator Protein 1 and its Ligands: Organising Chromatin Domains. Nucleic Acids and Molecular Biology *9*, 308-327.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., and Gasser, S. M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. J Mol Biol *231*, 293-310.

Goncalves, P. M., Maurer, K., van Nieuw Amerongen, G., Bergkamp-Steffens, K., Mager, W. H., and Planta, R. J. (1996). C-terminal domains of general regulatory factors Abf1p and Rap1p in Saccharomyces cerevisiae display functional similarity. Mol Microbiol *19*, 535-43.

Gotta, M., and Gasser, S. M. (1996). Nuclear organization

and transcriptional silencing in yeast. Experientia *52*, 1136-47.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., and Gasser, S. M. (1996). The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. J Cell Biol *134*, 1349-63.

Gotta, M., Strahl-Bolsinger, S., Renauld, H., Laroche, T., Kennedy, B. K., Grunstein, M., and Gasser, S. M. (1997). Localization of Sir2p: the nucleolus as a compartment for silent information regulators. Embo J *16*, 3243-55.

Gottschling, D. E. (1992). Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. Proc Natl Acad Sci U S A *89*, 4062-4065.

Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. Cell *63*, 751-762.

Graham, I. R., and Chambers, A. (1994). Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor. Nucleic Acids Res *22*, 124-30.

Hardy, C. F., Balderes, D., and Shore, D. (1992). Dissection of a carboxy-terminal region of the yeast regulatory protein RAP1 with effects on both transcriptional activation and silencing. Mol Cell Biol *12*, 1209-17.

Hardy, C. F., Sussel, L., and Shore, D. (1992). A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. Genes Dev *6*, 801-14.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell *80*,

583-92.

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. (1996). Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. Nature *383*, 92-6.

Henry, Y. A., Chambers, A., Tsang, J. S., Kingsman, A. J., and Kingsman, S. M. (1990). Characterisation of the DNA binding domain of the yeast RAP1 protein [published erratum appears in Nucleic Acids Res 1990 Jul 25;18(14):4317]. Nucleic Acids Res *18*, 2617-23.

Herskowitz, I. (1988). Life Cycle of the Budding Yeast Saccharomyces cerevisiae. Microbiological Reviews *52*, 536-553.

Hoffman, C. S., and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene *57*, 267-272.

Hofmann, J. F. X., Laroche, T., Brand, A. H., and Gasser, S. M. (1989). RAP-1 factor is necessary for DNA loop formation in vitro at the silent mating type locus HML. Cell *57*, 725-737.

Huet, J., Cottrelle, P., Cool, M., Vignais, M.-L., Thiele, D., Marck, C., Buhler, J.-M., and Sentenac, A. (1985). A general upstream binding factor for genes of the yeast translational apparatus. Embo J *4*, 3539-3547.

Johnson, L. M., Kayne, P. S., Kahn, E. S., and Grunstein, M. (1990). Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *87*, 6286-6290.

Kayne, P. S., Kim, U.-J., Han, M., Mullen, J., Yoshizaki, F., and Grunstein, M. (1988). Extremely Conserved Histone H4 N Terminus Is Dispensable for Growth but Essential for Repressing the Silent Mating Loci in Yeast. Cell *55*, 27-39.

Klein, C., and Struhl, K. (1994). Protein kinase A

mediates growth-regulated expression of yeast ribosomal protein genes by modulating RAP1 transcriptional activity. Mol Cell Biol *14*, 1920-8.

Klein, F., Laroche, T., Cardenas, M. E., Hofmann, J. F., Schweizer, D., and Gasser, S. M. (1992). Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. J Cell Biol *117*, 935-48.

Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell *85*, 125-36.

Kowalczykowski, S. C., Lonberg, N., Newport, J. W., von Hippel, P. H. (1981). Interactions of Bacteriophahe T4-coded Gene 32 Protein with Nucleic Acids: I. Characterization of the Binding Interactions. J Mol Biol *145*, 75-104.

Kraakman, L. S., Mager, W. H., Maurer, K. T., Nieuwint, R. T., and Planta, R. J. (1989). The divergently transcribed genes encoding yeast ribosomal proteins L46 and S24 are activated by shared RPG-boxes. Nucleic Acids Res *17*, 9693-706.

Kramer, H., Niemoller, M., Amouyal, M., Revet, B., Wilcken-Bergmann, v., and Muller-Hill, B. (1987). *lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators. Embo J *6*, 1481-1491.

Krauskopf, A., and Blackburn, E. H. (1996). Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats. Nature *383*, 354-7.

Kurtz, S., and Shore, D. (1991). RAP1 protein activates and silences transcription of mating-type genes in yeast. Genes Dev *5*, 616-28.

Kyrion, G., Boakye, K. A., and Lustig, A. J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in Saccharomyces cerevisiae. Mol Cell Biol *12*, 5159-73.

Kyrion, G., Liu, K., Liu, C., and Lustig, A. J. (1993). RAP1 and telomere structure regulate telomere position effects in Saccharomyces cerevisiae. Genes Dev *7*, 1146-59.

Larson, G. P., Castanotto, D., Rossi, J. J., and Malafa, M. P. (1994). Isolation and functional analysis of a Kluyveromyces lactis RAP1 homologue. Gene *150*, 35-41.

Laurenson, P., and Rine, J. (1992). Silencers, silencing and heritable transcription states. Microbiol. Rev. *56*, 543-560.

Liu, C., and Lustig, A. J. (1996). Genetic analysis of Rap1p/Sir3p interactions in telomeric and HML silencing in Saccharomyces cerevisiae. Genetics *143*, 81-93.

Liu, C., Mao, X., and Lustig, A. J. (1994). Mutational analysis defines a C-terminal tail domain of RAP1 essential for Telomeric silencing in Saccharomyces cerevisiae. Genetics *138*, 1025-40.

Longtine, M. S., Enomoto, S., Finstad, S. L., and Berman, J. (1993). Telomere-Mediated Plasmid Segregation in *Saccharomyces cerevisiae* Involves Gene Products Required for Transcriptional Repression at Silencers and Telomeres. Genetics *133*, 171-182.

Longtine, M. S., Wilson, N. M., Petracek, M. E., and Berman, J. (1989). A yeast telomere binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1. Curr Genet *16*, 225-39.

Loo, S., and Rine, J. (1994). Silencers and Domains of Generalized Repression. Science *264*, 1768-1771.

Lustig, A. J., Kurtz, S., and Shore, D. (1990). Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. Science *250*, 549-53.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Indust. Appl. Math. *11*, 431-441.

Moazed, D., and Johnson, A. D. (1996). A deubiquinating enzyme interacts with SIR4 and regulates silencing in *S. cerevisiae*. Cell *86*, 667-677.

Moazed, D., Kistler, A., Axelrod, A., Rine, J., and Johnson, A. D. (1997). Silent information regulator protein complexes in Saccharomyces cerevisiae: A SIR2/SIR4 complex and evidence for a regulatory domain in SIR4 that inhibits its interaction with SIR3. Proc Natl Acad Sci U S A *94*, 2186-2191.

Moretti, P., Freeman, K., Coodly, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes Dev *8*, 2257-69.

Muller, T., Gilson, E., Schmidt, R., Giraldo, R., Sogo, J., Gross, H., and Gasser, S. M. (1994). Imaging the asymmetrical DNA bend induced by repressor activator protein 1 with scanning tunneling microscopy. J Struct Biol *113*, 1-12.

Nasmyth, K. A. (1982). The Regulation of Yeast Mating-Type Chromatin Structure by SIR: An Action at a Distance Affecting Both Transcription and Transposition. Cell *30*, 567-578.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S. M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. Cell *75*, 543-55.

Retamal, C., and Babul, J. (1988). Determination of the Molecular Weight of Proteins by Electrophoresis in Slab Gels with a Transverse Pore Gradient of Crosslinked Polyacrylamide in the Absence of Denaturing Agents. Analytical Biochemistry *175*, 544-547.

Schleif, R. (1992). DNA Looping. Annual Review of Biochemistry *61*, 199-223.

Schultz, S. C., Shields, G. C., and Steitz, T. A. (1990). Crystallization of *Escherichia coli* Catabolite Gene Activator Protein with its DNA Binding Site: The use of

Modular DNA. J Mol Biol *213*, 159-166.

Senear, D. F., and Brenowitz, M. (1991). Determination of Binding Constants for Cooperative Site-specific Protein-DNA Interactions Using the Gel Mobility-shift Assay. J Biol Chem *266*, 13661-13671.

Senear, D. F., Brenowitz, M., Shea, M. A., and Ackers, G. K. (1986). Energetics of Cooperative Protein-DNA Interactions: Comparison between Quantitative Deoxyribonuclease Footprint Titration and Filter Binding. Biochemistry *25*, 7344-7354.

Shanmugan, K., Featherstone, M. S., and Saragovi, H. U. (1997). Residues Flanking the HOX YPWM motif contribute to cooperative interactions with PBX. J Biol Chem *272*, 19081-19087.

Shore, D. (1994). RAP1: a protean regulator in yeast. Trends Genet *10*, 408-12.

Shore, D., and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell *51*, 721-32.

Shore, D., Stillman, D. J., Brand, A. H., and Nasmyth, K. (1987). Identification of silencer bindng proteins from yeast: possible roles in SIR control and DNA replication. Embo J *6*, 461-467.

Sorger, P. K., and Pelham, H. R. B. (1988). Yeast Heat Shock Factor Is an Essential DNA-Binding Protein That Exhibits Temperature-Dependent Phosphorylation. Cell *54*, 855-864.

Sussel, L., and Shore, D. (1991). Separation of transcriptional activation and silencing functions of the RAP1-encoded repressor/activator protein 1: isolation of viable mutants affecting both silencing and telomere length. Proc Natl Acad Sci U S A *88*, 7749-53.

Tanaka, I., Appelt, K., Dij, K. L., White, S. W., and Wilson, K. S. (1984). 3 A resolution structure of a protein with histone-like properties in prokaryotes.

Nature *310*, 376-381.

Thompson, J. S., X., L., and Grunstein, M. (1994). The histone H3 amino terminus is required for both telomeric and silent mating locus repression in yeast. Nature *369*, 245-247.

Tornow, J., Zeng, X., Gao, W., and Santangelo, G. M. (1993). GCR1, a transcriptional activator in Saccharomyces cerevisiae, complexes with RAP1 and can function without its DNA binding domain. Embo J *12*, 2431-7.

Tsang, J. S., Henry, Y. A., Chambers, A., Kingsman, A. J., and Kingsman, S. M. (1990). Phosphorylation influences the binding of the yeast RAP1 protein to the upstream activating sequence of the PGK gene. Nucleic Acids Res *18*, 7331-7.

Vignais, M.-L., Huet, Janine, Buhler, J.-M., and Sentenac, A. (1989). Asymmetric DNA Bending Induced by the Yeast Multifunctional Factor TUF. J Biol Chem *264*, 8463-8466.

Vignais, M.-L., Huet, Janine, Buhler, J.-M., and Sentenac, A. (1990). Contacts between the Factor TUF and RPG Sequences. The Journal of Biological Chemistry *265*, 14669-14674.

Wang, S. S., and Zakian, V. A. (1990). Sequencing of *Saccaromyces* telomeres cloned using T4 DNA polymerase reveals two domains. Mol Cell Biol *10*, 4415-4419.

Warner, J. R. (1989). Synthesis of Ribosomes in *Saccharomyces cerevisiae*. Microbiological Reviews *53*, 256-271.

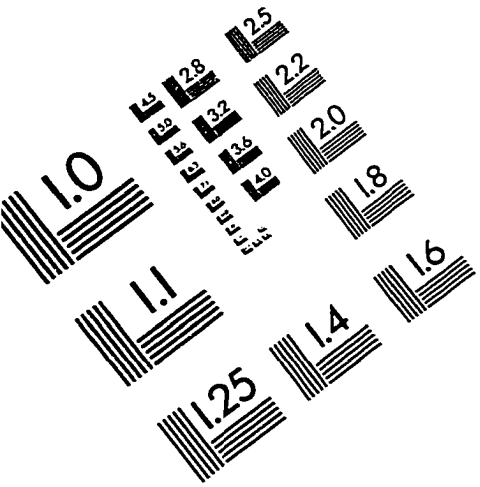
Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. Genes Dev *11*, 748-60.

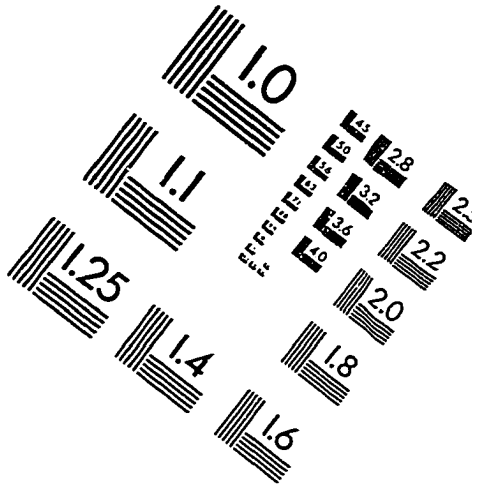
Wright, J. H., Gottschling, D. E., and Zakian, V. A. (1992). Saccharomyces telomeres assume a non-nucleosomal

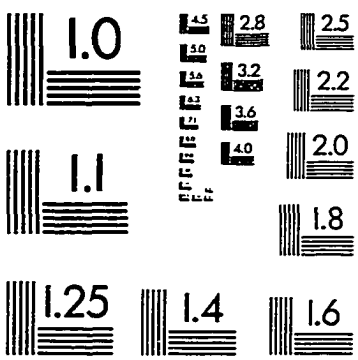
chromatin structure. Genes Dev *6*, 197-210.

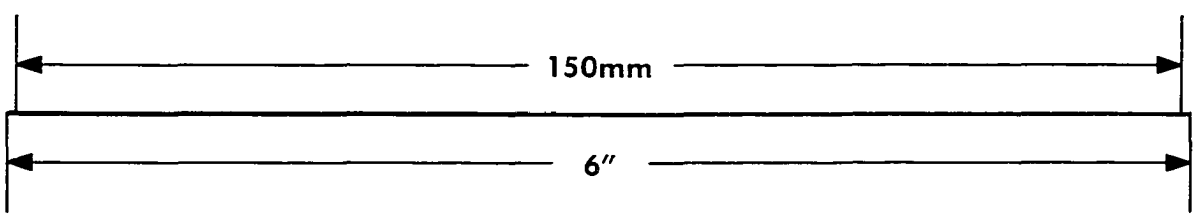
Wright, J. H., and Zakian, V. A. (1995). Protein-DNA interactions in soluble telosomes from *Saccharomyces cerevisiae*. Nucleic Acids Research *23*, 1454-1460.

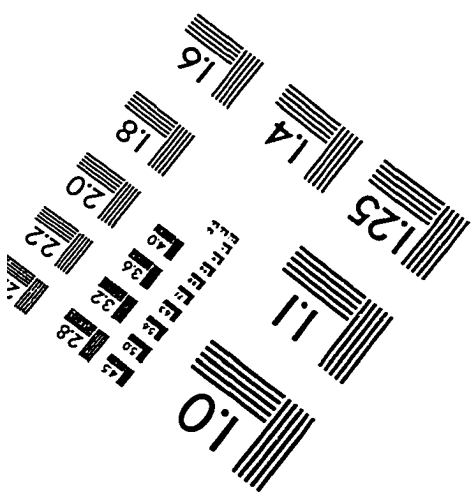
IMAGE EVALUATION TEST TARGET (QA-3)

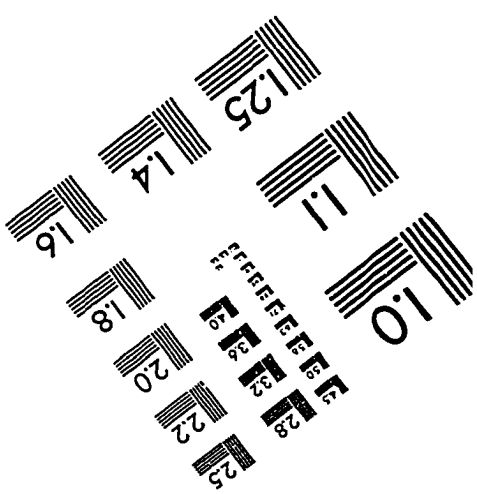












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

EXAMINATION OF VARIOUS TECHNIQUES FOR MEASURING WIND VELOCITIES USING CLEAR-AIR RADARS, WITH EMPHASIS ON VERTICAL WIND MEASUREMENTS

by

JORGE LUIS CHAU

B.S., University of Piura, Perú, 1991

M.S., University of Colorado, 1995

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

1998

This thesis for the Doctor of Philosophy degree by

Jorge Luis Chau

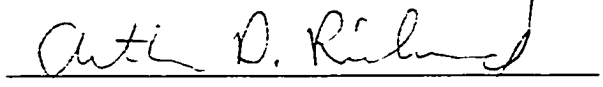
has been approved for the

Department of

Electrical and Computer Engineering

by

Ben B. Balsley



Arthur D. Richmond

Date January 16, 1998

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Chau. Jorge Luis (Ph. D., Electrical Engineering)

Examination of various techniques for measuring wind velocities using clear-air radars, with emphasis on vertical wind measurements

Thesis directed by Research Professor Ben B. Balsley

In this work, I examine various techniques for measuring wind velocities using clear-air radars. Special emphasis has been given to the vertical wind measurements. All the observations were obtained in the lower atmosphere (below 20 km) using the Jicamarca 50 MHz radar system near Lima, Perú. I have concentrated my efforts on: (a) examining different radar techniques for horizontal wind estimation, (b) improving the understanding of angle-of-arrival radar measurements, and (c) comparing a number of techniques for the measurement of the vertical wind velocities.

I have measured horizontal wind velocities by a variety of both time- and frequency-domain spaced antenna (SA) techniques. Comparisons of these techniques for both zonal and meridional components have been carried out in a statistical sense. Two sets of data were analyzed under both "quiet" (low wind, low variability) and "active" (high wind, high variability) conditions. I have found that the simpler techniques that assume horizontally isotropic scattering compare well with the more complicated full correlation analysis techniques. Briefly, while all SA techniques give essentially the same horizontal wind direction, considerable discrepancies are apparent in the wind speed, particularly above 15 km.

With regard to technique comparisons for vertical velocity, two types of measurement techniques using a concurrent data set are compared. Both Zenith (i.e., using a vertically-directed single beam) techniques and spaced antenna (SA) techniques are compared using a concurrent 5-day data set obtained with a special configuration of the Jicamarca 50 MHz radar antenna. Two separate Zenith

techniques are implemented using different antenna beam-widths (a 0.85° and a 3° beam-width). Different SA approaches are also implemented using a: "quasi" Doppler beam swinging (DBS) approach and four "classical" SA approaches. "Classical" SA approaches make use of horizontal wind and AOA measurements to get "corrected" vertical velocities. Statistical results are presented for both quiet and active conditions.

DEDICATION

To Jenny and Tiffany

ACKNOWLEDGMENTS

First of all, I would like to thank my lovely wife (Jenny) and daughter (Tiffany), without their constant support throughout these years the END of this thesis would have not been possible.

Special thanks to my advisor Ben Balsley for his continuous help not only in the scientific arena but also in personal matters. His simple way of facing the obstacles has contributed to ease my "complicated" way of work.

Ronald Woodman patiently taught me the "teje y maneje" of gathering experimental radar data. Discussions on how to implement the experiments and on how to process and interpret the data, has enormously contributed to the success of this research.

I would like to thank Susan Avery, Zoya Popovic, Art Richmond, Koni Steffen and Judy Curry for being members of my committee.

The Jicamarca staff is also acknowledged. The games of ping-pong and "fulbito" and the "historical" anecdotes contribute to a pleasant "hands-on" experience. The dedication and knowledge of Fernando Villanueva made possible the gathering of good and uninterrupted data (you still own me a "cebiche").

Thanks also to my officemates. Luis "the coolest one" Liziola for sharing his "cool" IDL routines, especially "make_x.pro". Meche Huaman (my "Comadre") for bringing "el toque femenino" to the office, for her time and support, and of course for her delicious "platillos". Mike Jensen for bringing the diversity to the group (Kites = Radars?).

Frequent and fruitful discussions with Peter May helped me to investigate better the SA techniques and AOA measurements. I also would like to thank Denise

Thorsen for reading and commenting on a portion of this work (AOA chapter) and Art Richmond for his careful reading of early manuscripts.

Brief but important discussions with the following scientists are greatly appreciated: Erhan Kudeki (narrow-wide beam idea. SA implementation). Jürgen Röttger (specular echoes and AOA measurements). Dick Doviak (scattering theory. AOA determination). Richard Lataitis (SA techniques). Christopher Holloway (SA techniques). Greg Nastrom (vertical velocity issues). Robert Palmer (vertical velocity issues). Mike Jones (gravity waves). Rod Frehlich (statistical uncertainties).

Finally, I would like to thank my parents for their support and time in the final stages of this "journey".

The work in this thesis was funded through grants from the National Science Foundation under agreements ATM-9214657 and ATM-9614700.

CONTENTS

CHAPTER

1 INTRODUCTION 1

2 CLEAR-AIR VERTICAL MOTIONS 3

2.1 On the importance of knowing vertical atmospheric motions . . 4

2.1.1 The importance of vertical air motions in meteorology 4

2.1.2 The importance of vertical air motions in atmospheric research 5

2.1.3 The importance of vertical air motions in air pollution 6

2.1.4 The importance of vertical air motions in modeling 6

2.1.5 The importance of vertical air motions in horizontal wind estimation 7

2.2 Methods of Estimating the Vertical Wind Velocity 7

2.2.1 Non-radar methods 7

2.2.1.1 Adiabatic method 8

2.2.1.2 Kinematic method 8

2.2.1.3 Methods using Quasi-Geostrophic theory 9

2.2.2 Clear-air radar methods 9

2.2.2.1 The VAD (Velocity Azimuth Display) technique 10

2.2.2.2 The Zenith technique 11

2.2.2.3 Spaced antenna techniques 11

2.2.2.4 One form of the adiabatic method 12

2.3 Vertical Velocity Characteristics 13

2.3.1 Vertical velocity characteristics from theory 13

2.3.2 Some observed vertical velocity characteristics 15

2.3.2.1 Mean vertical motions 15
2.3.2.2 The variability of vertical velocity 16
2.4 Current concerns with the Zenith technique 18
2.4.1 Antenna beam pointing inaccuracies 20
2.4.2 Tilted refractive index structures 20
2.4.3 Gravity waves 21
2.4.4 Wind field gradients 21
3 FOCUSING IN ON THE VERTICAL VELOCITY PROBLEM . . . 23
4 A STATISTICAL COMPARISON OF HORIZONTAL WINDS OBTAINED BY A VARIETY OF SPACED ANTENNA TECHNIQUES USING THE JICAMARCA VHF RADAR 26
4.1 Introduction 26
4.2 Spaced-Antenna Techniques 28
4.2.1 Time-Domain techniques 29
4.2.2 Frequency-Domain techniques 32
4.3 Jicamarca Experiment 34
4.4 Results 37
4.4.1 Quiet period results 38
4.4.2 Active period results 43
4.5 Discussion 47
4.6 Conclusions 54
5 INTERPRETATION OF ANGLE-OF-ARRIVAL (AOA) MEASUREMENTS IN THE LOWER ATMOSPHERE USING SPACED ANTENNA RADAR SYSTEMS 57
5.1 Introduction 57
5.2 Angle-of-arrival theory for SA systems 59
5.2.1 AOA contributions 59

5.2.2 Methods of estimating AOAs 65
5.2.3 Statistical errors in time-domain AOA measurements 66
5.3 On the use of non-collocated SA systems for AOA measurements 67
5.4 Experimental results 69
5.5 Conclusions 79
6 A STATISTICAL COMPARISON OF VHF TECHNIQUES TO STUDY CLEAR-AIR VERTICAL VELOCITIES IN THE LOWER ATMOSPHERE USING THE JICAMARCA RADAR 82
6.1 Introduction 82
6.2 VHF radar techniques for vertical wind velocity measurements . 85
6.2.1 The Zenith technique 85
6.2.2 Spaced-antenna (SA) techniques 86
6.2.2.1 The "quasi" DBS approach 86
6.2.2.2 The "classical" SA approach 87
6.3 Experimental setup 88
6.4 Results 91
6.5 Discussion 105
6.6 Conclusions 110
7 SUMMARY AND FUTURE WORK 113
BIBLIOGRAPHY 118
A GROUND-CLUTTER TREATMENT 128
A.1 Time-Domain subtraction 129
A.2 Frequency-Domain subtraction 130
A.3 Frequency-Domain interpolation 132
APPENDIX
B SPECTRAL MOMENT ESTIMATION 133
B.1 NOAA's algorithm 133

B.2 Gaussian fitting algorithm 135

C PHASE CALIBRATION OF SPACED ANTENNA SYSTEMS 137

D IMPLEMENTATION OF THE POST STATISTICS STEERING TECHNIQUE 139

E AOA INTERPRETATION UNDER HORIZONTALLY ANISOTROPIC SCATTERING 142

FIGURES

FIGURE

4.1 Antenna setup for the SA experiment at Jicamarca. The transmitting antenna is the D module, pointing vertically; and the receiving antennas are A,B and C modules. Notice that neither of the antenna axe, (x, y), are aligned with the geographic coordinates (E, N). Moreover, the antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest. 36

4.2 Mean horizontal velocity profiles from all techniques for the quiet period (20-Jan-97 to 27-Jan-97). (a) Magnitude. (b) Direction and (c) Mean SNR profile (the vertical dashed line represents the SNR threshold used to reject data). Relative magnitude difference in percent (Δ) for (d) time-domain techniques and (e) frequency-domain techniques at the bottom. See Section 4.2 for description of the different techniques. 39

4.3 Examples of correlation and histogram comparisons of 3-minute zonal velocities obtained by the FCAMt and the FSA SA techniques during the quiet period. (a) and (b) High SNR (5-10 km), and (c) and (d) Low SNR (16-18 km) comparisons. ρ_d represents the correlation coefficient. μ_d the mean and σ_d the standard deviation of the velocity differences (FCAMt - FSA). 41

4.4 Statistical comparisons of hourly-average zonal velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray): during the quiet period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences. 42

4.5 Statistical comparisons of hourly-average meridional velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray); during the quiet period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences. 43

4.6 Mean horizontal velocity profiles from all techniques for the active period (25-Mar-97 to 29-Mar-97). (a) Magnitude, (b) Direction and (c) Mean SNR profile (the vertical dashed line represents the SNR threshold used to reject data). Relative magnitude difference in percent (Δ) for (d) time-domain techniques and (e) frequency-domain techniques at the bottom. See Section 4.2 for description of the different techniques. 45

4.7 Examples of correlation and histogram comparisons of 3-minute zonal velocities obtained by the FCAMt and the FSA SA techniques during the active period. (a) and (b) High SNR (5-10 km). and (c) and (d) Low SNR (16-18 km) comparisons. ρ_d represents the correlation coefficient. μ_d the mean and σ_d the standard deviation of the velocity differences (FCAMt - FSA). 46

4.8 Statistical comparisons of hourly-average zonal velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray); during the active period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences. 47

4.9 Statistical comparisons of hourly-average meridional velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray); during the active period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences. 48

4.10 Percentage of 3-minute zonal velocity failures (α) for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray). (a) The quiet period. (b) The active period. This failures represents the estimates rejected in both cleaning stages: "pre-cleaning" and "post-cleaning". 49

5.1 Antenna setup for the Jicamarca experiment. The transmitting antenna is the vertically pointing D module; and the receiving antennas are the "on-axis" pointing A,B and C modules. The "on-axis" position differ from the zenith position because antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest. Notice that neither of the antenna axes, (x, y), is aligned with the geographic coordinates (E, N). The antenna centers are given in (x, y) coordinates, respect to the transmitting antenna center. 71

5.2 Histograms of 3-minute AOAs obtained by a time-domain method (see Section 5.2.2) using the BC (Θ_x), and BA (Θ_y) antenna pairs. Solid lines represent the actual measurements, and the fitted Gaussian functions are represented by stars. Mean (μ_Θ) and standard deviation (σ_Θ) of the AOA measurements are given for the "raw" (subscript r) and Gaussian fitted (subscript g) data. These AOA measurements are relative to the antenna plane (see Figure 5.1) and include heights between 7 and 8 km. 73

5.3 Histograms of 3-minute AOAs obtained by a frequency-domain method (see Section 5.2.2) using the BC (Θ_x), and BA (Θ_y) antenna pairs. Solid lines represent the actual measurements, and the fitted Gaussian functions are represented by stars. Mean (μ_Θ) and standard deviation (σ_Θ) of the AOA measurements are given for the "raw" (subscript r) and Gaussian fitted (subscript g) data. These AOA measurements are relative to the antenna plane (see Figure 5.1) and include heights between 7-8 km. 74

5.4 Theoretical AOAs under isotropic conditions ($\xi_h^2 \approx 2a_h^{-2}$), including just the SA system contributions. Transmitting beam contribution ($2\Psi\xi_h^{-2}$), two geometric terms ($2\Gamma\xi_h^{-2}$ and Λ), and the combined contribution $[2(\Psi+\Gamma)\xi_h^{-2}+\Lambda]$. (a) BC ($\Theta_x$), and (b) BA ($\Theta_y$). 75

5.5 AOA profiles for the five-day data set. AOA results for (a) baseline BC (Θ_x) and (b) baseline BA (Θ_y). The mean SNR profile appears in (c). The means and standard deviations of empirical values (triangles and horizontal dotted lines, respectively), in panels (a) and (b) using the bottom scales, along with the mean theoretical value for isotropic conditions (thick solid line). Statistical errors (ε_Θ) denoted by diamonds are shown in panels (a) and (b) plotted using the top scales. 76

5.6 Examples of AOA autocorrelation functions ((a) and (c)) and AOA time-series ((b) and (d)) for baselines BC (Θ_x), and BA (Θ_y) respectively. AOA values for the 3-minute measurements are represented in (b) and (d) by stars. Corresponding 15-minute smoothed curves are shown by a thick continuous line. The empirical statistical variances are denoted by $\hat{\varepsilon}^2_\Theta$ (the horizontal lines in (a) and (c)), and the theoretical value by ε^2_Θ (see Equation 5.29). The empirical values are obtained from the difference between the autocorrelation of the 3-minute time-series and the autocorrelation of the smoothed time-series, at $\tau = 0$. 77

5.7 Histograms of (a) correlation times (τ_c), and coherence for baselines (b) BC ($\mathrm{coh}_{BC}[0]$) and (c) BA ($\mathrm{coh}_{BA}[0]$). The mean values in all panels are represented with vertical dashed lines, and the theoretical values under isotropic conditions ($\xi^2_h \approx 2a_h^{-2}$), panels (b) and (c), by vertical solid lines. The anlysis is done for heights between 6.15 to 8.40 km. 78

5.8 Time series of 3-minute AOAs for six tropospheric heights (6.15 to 8.40 km). Results for baseline (a) BC (Θ_x), and (b) BA (Θ_y). Mean theoretical values under isotropic conditions ($\xi_h^2 \approx 2a_h^{-2}$) represented by long dashed lines. The short-dashed lines in panel (b) represent the zero values at each height. Notice the long-period oscillations (> 12 hours). 80

6.1 Antenna setup for the Jicamarca experiment. The full array, vertically pointing E module, was used for transmission and reception in the "Down" polarization (x-aligned). In the orthogonal polarization ("Up", y-aligned), the vertically pointing D module was used for transmission and reception. The "on-axis" pointing A,B and C modules are used only for reception. The "on-axis" position differ from the zenith position, because the antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest. Notice that neither of the antenna axes, (x, y), is aligned with the geographic coordinates (E, N). 90

6.2 Mean SNR profiles for the following time periods: (a) All (b) Quiet and (c) Active. The SA $\overline{\mathrm{SNR}}$ profile is denoted by "+s", while the wide beam and narrow beam $\overline{\mathrm{SNR}}$ are denoted with asterisks (*) and diamonds ($\diamond$) respectively. 93

6.3 (a) Median turbulence intensity (σ_t), and (b) median correlation length (ξ'), for the time-periods: All (+), Quiet (*) and Active ($\diamond$). 94

6.4 Profiles of mean zonal and meridional wind velocities (μ_u and μ_v) for the following SA techniques: the reference profile FCAM (+), FCAB (*), FCAH ($\diamond$), FSAM (Δ), and PSS ($\Box$). Each column corresponds to a particular selected time period (All: first column, Quiet: second column, and Active: third column). 95

6.5 Zonal and meridional AOAs for the three selected time-periods: All (first column), Quiet (second column) and Active (third column) time periods. Median values are shown by diamonds and standard deviations are shown by horizontal solid lines. The solid curve in each plot represents theoretical values from Chapter 5 expected from an isotropic atmosphere, given the antenna geometry and pointing direction. The vertical dashed lines represent AOAs of zero degrees. 97

6.6 Profiles of vertical velocity characteristics: standard deviations (σ_w, upper panels) and percentage of failures (α_w, lower panels), for All (first column), Quiet (second column) and Active (third column) time periods. The w measurements are: Wn (+), FCAM (*), PSS ($\diamond$) and Ww ($\triangle$). 98

6.7 Profiles of vertical velocity comparisons for the following FCA-type measurements: FCAB (*), FCAH ($\diamond$), FCAM ($\triangle$) and FSAM ($\square$), using Wn as a reference. The comparisons are in terms of: mean velocity differences ($\mu_{\Delta w}$, first column), standard deviation of the velocity differences ($\sigma_{\Delta w}$, second column) and correlation coefficient ($\rho_{\Delta w}$, third column), for All (first row), Quiet (second row) and Active (third row). 100

6.8 Profiles of vertical velocity comparisons for the following measurements: FCAr (*), PSS ($\diamond$) and Ww ($\triangle$), using Wn as a reference. The comparisons are in terms of: mean velocity differences ($\mu_{\Delta w}$, first column), standard deviation of the velocity differences ($\sigma_{\Delta w}$, second column) and correlation coefficient ($\rho_{\Delta w}$, third column), for All (first row), Quiet (second row) and Active (third row). . . 101

6.9 Profiles of mean vertical velocities (μ_w) for time-intervals of events (a) All, (b) Quiet and (c) Active. The measurements are: the reference profile Wn(+), FCAM (*), PSS ($\diamond$), and Ww ($\triangle$). 103

6.10 Standard deviations of the differences ($\sigma_{\Delta w}$ by stars in upper panels) and correlation coefficients ($\rho_{\Delta w}$ by stars in lower panels). In addition to the standard "errors" of vertical velocities (σ_r by diamonds in first column) and theoretical-derived correlation coefficients ($\hat{\rho}_{\Delta w}$). 104

A.1 Real part of a time-series for 10 blocks of 64 points each. Solid circles denote the mean values of each block, and vertical dotted lines delimit each block. The thick line represents the fitted second-degree polynomial to the circles. The ground-clutter-free signal is at the bottom, centered on zero. 129

A.2 Example of atmospheric power spectrum with ground-clutter and noise (thick line). Ground-clutter spectrum is denoted with pluses and the ground-clutter-free spectrum with triangles. 131

B.1 An example of a frequency spectrum represented by Equation B.1. 134

B.2 Block diagram for a least squares parameter estimation. 136

TABLES

TABLE

4.1 SA technique abbreviations and some characteristics. 35

4.2 Operating parameters for the experiments conducted at Jicamarca 37

4.3 β_h results from histograms of the calculated values using Equation 4.6. The theoretical value is 0.040 50

5.1 Operating parameters for the experiments conducted at Jicamarca (from 25-Mar-97 to 29-Mar-97) 70

6.1 Salient features of the SA techniques for horizontal wind measurements used in vertical wind velocity estimation. 88

6.2 Antennas characteristics. 89

6.3 Operating parameters for the experiments conducted at Jicamarca (from 25-Mar-97 to 29-Mar-97). 91

6.4 Information on the periods (All, Quiet and Active) used in these analyses. 92

7.1 Subjective ratings for vertical velocity techniques. [Wn: narrow beam Zenith, Ww: wide beam Zenith, PSS: "quasi" DBS, and $\hat{w}$: "classical" SA] 115

C.1 Phase calibration values for the 5-day experiment run in March 1997. 138

D.1 Effective beam pointing directions at 7.50 km (in degrees). 141

CHAPTER 1

INTRODUCTION

Atmospheric vertical motions are very important in predicting and understanding atmospheric processes that are responsible for our weather and climate. For example, it is fairly common knowledge that the upward displacement of a mass of saturated air will produce precipitation. However, vertical motions are usually so small that when meteorologists use the term "wind" they usually are referring to the horizontal motions. This usage is in line with the Webster's dictionary definition of wind as: "A natural motion of the air: especially, a noticeable current of air moving in the atmosphere parallel to the earth's surface."

Throughout this thesis, we will be discussing and comparing atmospheric results obtained by VHF clear-air radars in the lower atmosphere (below 22 km), where most meteorological processes are confined. Clear-air radars offer the unique capability of directly measuring vertical motions. However, vertical motions obtained by these radars are not incorporated into numerical weather prediction (NWP) models: rather the vertical velocity is a derived field (Dunn [1991]). One can expect model performances to improve as clear-air radar measurements of the vertical velocity become more widely available to, and accepted by, the meteorological community. In contrast, horizontal winds measured by clear-air radars are more accepted and are being incorporated in meteorological models. For example, wind observations at Christmas Island (central Pacific) are being used routinely by the National Meteorological Center (NMC) and the European Centre for Medium Range Weather Forecasts (ECMWF) in their analysis and forecast products (Gage et al. [1988]).

The term clear-air radar can be misleading; although radar studies of the

atmosphere via scattering from clear-air refractive index fluctuations are the norm. VHF radars are also sensitive to strong precipitation (e.g. Currier et al. [1992]; May and Rajopadhyaya [1996]). These types of radars are also called: (a) wind profilers, because the measurement of vertical profiles of the three-dimensional wind is their main application, or (b) ST radars for the height range they cover (lower stratosphere and troposphere). However, clear-air radars are not limited to meteorological applications. They are also used in middle atmospheric and ionospheric studies. Moreover, the first wind measurements by these type of radars were done at the ionospheric facility (at that time) of Jicamarca, Perú in 1974 (Woodman and Guillén [1974]). A comprehensive history of clear-air radar studies can be found in Hardy and Gage [1990].

Given the meteorological community's skepticism of radar measured vertical velocities, this thesis represents a concentrated effort to examine different ways atmospheric vertical motions can be obtained by clear-air radars. The core of this work, i.e., the comparisons of vertical velocity techniques, is covered in Chapter 6. In order to better understand vertical motions, we first present some of their important aspects in Chapter 2. In this chapter, we outline their importance, methods for measuring them and some of their characteristics (from theory and from observations). In addition, we summarized the current concerns with vertical velocities measured by clear-air radars. Then, we focus in on the vertical velocity problem in Chapter 3. In Chapter 2, we also see the importance of accurate horizontal wind velocity and angle-of-arrival measurements, in a specific type of vertical velocity technique. Since these techniques could offer new insights into the vertical velocity accuracy, we have dedicated Chapter 4 to compare a variety of techniques that measure horizontal wind velocities (zonal [u] and meridional [v] components), and Chapter 5 to improve the understanding of angle-of-arrival measurements. Finally, a summary of results and future activities is outlined in Chapter 7.

CHAPTER 2

CLEAR-AIR VERTICAL MOTIONS

The vertical velocities we study throughout this thesis are those obtained by VHF radars under clear-air conditions. Obviously, clear air does not contain such awesome phenomena as tornados or baseball-sized hail found in severe storms. Clear-air structure, nevertheless, is rich in meteorological events that can lead to the development of these storms.

Much of our current knowledge of the role of vertical velocity in weather and climate is a result of the observational analysis of large-scale synoptic wind fields obtained using non-radar methods (see Section 2.2.1). Such analyses are highly smoothed and subject to uncertainties, and give us no information about temporal and spatial variability on smaller scales. Local vertical velocities, on the other hand, contain contributions from convection, waves, turbulence. Instantaneous vertical velocities can be as large as several meters per seconds. These local influences (e.g., orography) are likely to dominate the much weaker synoptic-scale vertical motions, which are typically in the range of a few centimeters per second.

In this section we summarize the importance of knowing the vertical motions of clear air. We also provide a review of the current methods of inferring or directly-measuring this quantity. We include some characteristics of the vertical wind obtained from theory and from observations, along with a review of the current concerns of the accuracy of direct radar measurements of the vertical velocity. Our summary covers synoptic-scale (days to weeks, 100 to 500 km) and mesoscale (minutes to days, 1 to 100 km) vertical motions.

2.1 On the importance of knowing vertical atmospheric motions

In the the following paragraphs we discuss the importance of vertical motions in: meteorology, atmospheric research, air pollution, modeling and horizontal wind estimation.

2.1.1 The importance of vertical air motions in meteorology

Rising air is typically associated with cloudy conditions and precipitation. Subsiding air, on the other hand, produces adiabatic heating and clearing conditions. Because descending air is compressed and warmed, cloud formation and precipitation are unlikely in an anticyclone and "fair" weather can usually be expected with the approach of a high pressure system. Cyclones and "stormy" weather, on the other hand, can be expected with the approach of a low pressure system.

Vertical air motion is one of the most important variables for many forecasting applications and much of synoptic meteorology is involved with the prediction of vertical velocity. Many forecasting "rules of thumb" have their roots in the dynamical equations for vertical motion (e.g. strong positive vorticity advection equates to upward vertical motion). If a forecaster could monitor the vertical velocity field in real time, his predictions of severe weather would be improved. It is within the zero-to-six-hour time frame where credible forecasts of severe weather have the greatest impact on protecting life and property, and it is precisely within this time frame that continuous observations of the vertical velocity can best help weather forecasting. Given continuously available vertical velocity profiles, a forecaster could better monitor storm movement and development in real time (the now-casting approach).

Although mesoscale vertical motion (e.g., motions due to cold fronts and baroclinic storm systems) plays a significant role in the dynamics of synoptic-scale systems, it is not resolved by conventional data analysis (radiosonde and forecasting tools).

It may be possible to deduce horizontal divergence from the vertical profile of vertical velocity at a single station. This idea relies upon the relationship between horizontal divergence and the vertical gradient of vertical velocity for an incompressible fluid, $\nabla_H \cdot \boldsymbol{u} = -\partial w/\partial z$. Thus, to the extent that a representative profile of vertical motions is obtainable at a single station, it would be possible to infer the vertical divergence profile from the vertical velocity profile (Gage [1990]). Horizontal divergence is a very important dynamical quantity in weather forecasting.

2.1.2 The importance of vertical air motions in atmospheric research

A partial listing of atmospheric research topics that involves vertical motions follows:

• Studies of orographic effects. Vertical velocity measurements from clear-air radars show that the morphology of atmospheric activity in the troposphere and lower stratosphere correlates well with the presence/absence of major mountain ranges. Examples of these studies (Ecklund et al. [1985]; Balsley and Carter [1989]; Liziola and Balsley [1997]; Nastrom and VanZandt [1996]) include the generation of mountain waves, the propagation of these waves into the upper atmosphere under specific mean conditions, the retardation/acceleration of waves by wave breakdown, and the detrimental effects of atmospheric gravity waves on the accuracy of wind measurements.

• Knowledge of the vertical motions in mesoscale convective systems (MCSs) is important for understanding the physical processes operating within these systems (e.g. precipitation production, stratiform and convective region dynamics). One also needs to understand how these features interact with the large-scale environment through the transport of heat, moisture and momentum (e.g. Houze [1989]).

• Knowledge of the vertical wind field is of considerable importance in understanding the troposphere-stratosphere exchange processes crucial for climate change processes.

• In tropical regions, vertical velocity information helps to understand the important scale interactions that take place between convection and large scale circulations (e.g., Walker circulation). For example, Gage et al. [1991a] presented an example of a possible correlation of El Niño events with measurements of mean vertical motions.
• In atmospheric wave studies, vertical velocity measurements are important in: (a) quantifying large-scale transport and thermal signatures of wave-driven residual circulation at both lower and upper levels, (b) identifying the magnitude, nature, and spectral character of various wave and turbulence processes, and (c) assessing wave field anisotropy and forcing via vertical fluxes of horizontal momentum.

2.1.3 The importance of vertical air motions in air pollution

Many extensive and long-lived air pollution events can be linked directly to temperature inversions that develop in regions of sinking air ($w < 0$) associated with slow-moving anticyclonic centers of high air pressure.

Moreover, at night time, when incoming solar radiation is smallest, low level temperature inversions tend to prevail. If such conditions are accompanied by light winds, then diffusion rates will be particularly small and pollutants emitted from ground sources will accumulate in the inversion layer.

2.1.4 The importance of vertical air motions in modeling

On an extended scale, both vertical velocity and horizontal velocity data from a network of clear-air radars could be used in the initialization and updating of regional National weather prediction (NWP) models. Vertical velocity measurements are also important for the verification of numerical-model-derived kinematic, thermodynamic and micro-physical fields.

For example, given measurements of the synoptic-scale vertical velocity, a single station could provide time histories of the vertical motion that exist within baroclinic storm systems. This capability would permit more realistic descriptive

models to be formulated.

Modeling studies presented by Starr and Cox [1985] showed ascent on the mesoscale to synoptic-scale is quite important to the evolution of macroscopic cloud properties. Macroscopic characteristics such as cloud thickness. aerial coverage and optical depth have been shown to be quite sensitive to the large-scale vertical motion.

2.1.5 The importance of vertical air motions in horizontal wind estimation

At times and places where vertical motions are appreciable. vertical velocity measurements are necessary in order to obtain more accurate measurements of horizontal winds using clear-air radars. For example. Weber et al. [1992] concluded that. in the presence of large vertical velocity disturbances. hourly averaged horizontal winds can be sometimes be less meaningful than the measurements made more frequently at intervals of a few minutes.

2.2 Methods of Estimating the Vertical Wind Velocity

The methods of estimating clear-air vertical motions can be classified conveniently into: non-radar and clear-air radar methods. These methods are described separately in the following paragraphs.

2.2.1 Non-radar methods

Non-radar methods for deriving the vertical motion. w. are those used typically by the meteorological community. Such methods involve large-scale measurements, and provide indirect estimates from synoptic analysis. Intercomparisons of these standard methods began half a century ago (Panofsky [1946]) and are a continuous topic of research (Durran and Snellman [1987]: Dunn [1991]).

Non-radar methods of vertical velocity estimation are usually applied using isobaric coordinate systems, i.e., using pressure (p) as the vertical coordinate instead

of height (z), so that $\omega(p) = \mathrm{D}p/\mathrm{D}t$ is inferred rather than $w(z) = \mathrm{D}z/\mathrm{D}t$. These two vertical velocity quantities are related through.

$$\omega = -\rho g w \tag{2.1}$$

where ρ is the air density and g is the acceleration due to gravity. This expression may not be a good approximation when: (a) the ageostrophic wind component is relatively large (e.g. as in most boundary layer events): (b) the local pressure tendency is relatively large (e.g. in an intensifying system): (c) the atmosphere is not in hydrostatic equilibrium (e.g. in a thunderstorm). Some of the most common non-radar methods of estimating w are outlined below.

2.2.1.1 Adiabatic method

This method is based on the assumption that changes of state of atmospheric air are adiabatic. Techniques that use this method are: the advective technique. the single-station technique. the graphical technique and the isentropic technique. The adiabatic method is inaccurate in situations where strong diabatic heating is present. i.e.. in storms in which heavy rainfall occurs over a large area (e.g. Panofsky [1946]: Holton [1992]: Bluestein [1992]). One disadvantage of this method is that determination of the local rate of change of temperature ($\partial T/\partial t$) is required at very close intervals.

2.2.1.2 Kinematic method

This method depends on the principle of conservation of mass as expressed by the equation of continuity (e.g. Panofsky [1946]: Pedder [1981]). Application of the kinematic method requires knowledge of the horizontal divergence of the horizontal velocity. In order to determine this divergence the partial derivatives are generally estimated from the fields of zonal and meridional velocities by using finite-difference approximations. However, a 10% error in evaluating one of the horizontal

wind components can easily cause the estimated divergence to be in error by 100%. Despite the inaccuracy of the divergence measurements, kinematic computations of w profiles based upon unsmoothed wind data do contain useful information. This is because divergence errors are vertically correlated. Synoptic-scale vertical motions computed kinematically and adjusted by the O'Brien scheme (O'Brien [1970]) have been shown to correlate well with cloud and precipitation patterns (Bluestein [1992]).

2.2.1.3 Methods using Quasi-Geostrophic theory

These methods of diagnosing the vertical velocity requires only information on the geopotential distributions at a single time, and utilizes both the vorticity equation and the thermodynamic equation. Quasi-geostrophic methods make use of the omega equation (e.g Dunn [1991]). The omega equation unlike the continuity equation, provides a method of estimating w that does not depend on observations of the ageostrophic wind. In fact, direct wind observations are not required at all, nor does the omega equation require information on the temperature tendency, as required in the adiabatic method. Accurately estimating the terms needed in these methods from noisy observational data can be quite difficult (e.g. Holton [1992]; Barnes [1986]; Durran and Snellman [1987]).

2.2.2 Clear-air radar methods

Clear-air radars typically observe weak backscatter from half-wavelength components of turbulent inhomogeneities in the atmospheric radio refractive index (e.g. Woodman and Guillén [1974]). At low tropospheric altitudes the radar refractive index is usually dominated by the humidity term. Above mid-tropospheric altitudes, the refractive index is dominated by turbulence in the thermal field. The intensity of backscattered power depends upon both the magnitude of the background gradient of refractive index as well as the intensity of the turbulence. Sometimes the spatial distribution of radar refractive index in clear-air can be associated

with meteorological phenomena such as waves, turbulent layers, fronts, etc. The most important radar techniques that measure vertical velocities of these "invisible" targets are described below.

2.2.2.1 The VAD (Velocity Azimuth Display) technique

In the VAD technique, the zenith angle (θ) is fixed and the antenna is swept in azimuth at a constant rate (e.g. Peterson and Balsley [1979]). Data at many altitudes are obtained essentially simultaneously as the radar antenna is swept 360° in azimuth. For a homogeneous and time-invariant wind field within the field of view at a given altitude, the radial velocity so obtained will vary sinusoidally with azimuth. The amplitude of the sine curve is equal to the horizontal wind speed and the phase is related to the wind direction. In addition, the offset of the sinusoid from zero velocity (C_0) is proportional to the vertical velocity ($C_0 = w\cos\theta$).

When divergence of the horizontal wind is important, the offset C_0 becomes

$$C_0 = \left[w + \frac{z_0}{2}\tan^2\theta(\nabla_h \cdot u)\right]\cos\theta \tag{2.2}$$

where $\nabla_h \cdot u$ is the divergence term and z_0 is the height range (Doviak and Zrnić [1993], p. 312). The divergence and vertical velocity can be calculated directly if VAD measurements are made at two or more different zenith angles. This modified VAD approach is known as EVAD (extended VAD, e.g. Larsen et al. [1991]).

Vincent and Reid [1983] suggested a variation to this technique, using at least two coplanar beams pointed symmetrically relative to vertical. Measuring the radial velocities at angles θ and $-\theta$, and assuming the mean horizontal and vertical velocities are the same in both beams, the mean vertical velocity at height range z_0 can be obtained from

$$\bar{w}(z_0) = \frac{\bar{v}_r(\theta, z_0) + \bar{v}_r(-\theta, z_0)}{2\cos\theta} \tag{2.3}$$

where $\bar{v}_r$ represents the mean radial velocity at the respective beams.

2.2.2.2 The Zenith technique

The Zenith technique can be thought of as the vertical portion of the Doppler beam swinging (DBS) method. The DBS method is the radar method currently used by most clear-air radars. In the Zenith technique the antenna remains fixed at the zenith. Since the radial velocity measured in this mode is the vertical velocity, one need to make no assumptions regarding homogeneity, particularly if narrow beam antennas are used. Current concerns on the accuracy of Zenith techniques are covered in Section 2.4.

2.2.2.3 Spaced antenna techniques

All spaced antenna techniques require at least three non-collinear, vertically-directed receiving antennas. The vertical velocity is obtained from

$$v_r = u \sin \Theta_u + v \sin \Theta_v + w(1 - \sin^2 \Theta_u - \sin^2 \Theta_v)^{1/2} \tag{2.4}$$

where u, v are the horizontal velocity components (zonal and meridional), $\sin \Theta_u$, $\sin \Theta_v$ are the zonal and meridional direction cosines of the "effective" beam direction, and v_r is the radial velocity (positive going away from the radar). If these angles are measured from the phase of the cross-spectra functions (e.g. Van Baelen and Richmond [1991]; Palmer et al. [1991]; Larsen et al. [1992]), or from cross-correlation functions (e.g. Röttger and Ierkic [1985]; Larsen and Röttger [1991]), the vertical velocity obtained is often called corrected velocity (Larsen and Röttger [1991]). However, if different beam directions are obtained with the SA system, through post beam set (PBS, Röttger and Ierkic [1985]) or through post statistics steering (PSS, Kudeki and Woodman [1990]), the vertical velocity can be inferred from Equation 2.4. More on these techniques is presented in Section 6.2.

2.2.2.4 One form of the adiabatic method

One form of the adiabatic method described in Section 2.2.1.1 above, can be used with wind and temperature data from a single radar station. The station needs to have both a clear-air radar and a radio acoustic sounding system (RASS) (Nastrom and Warnock [1994]; Yamamoto et al. [1996]). With this type of station, wind and temperature measurements are available with the necessary temporal sequence to use this method on a routine basis. For example, Yamamoto et al. [1996] have shown preliminary measurements of vertical velocities obtained with the MU radar-RASS system. Assuming hydrostatic equilibrium, Yamamoto et al. obtained the vertical velocity from

$$w \approx -\frac{g}{N^2}\frac{\partial \ln T}{\partial t} \tag{2.5}$$

where g is gravity, T is temperature and N is the Brunt-Väisälä frequency. Yamamoto et al. found that "frequency spectra of the derived vertical wind were more consistent with a gravity wave model than those spectra calculated from direct measurements of the vertical wind".

In addition, some of the discussed non-radar methods could be applied to a network of clear-air radars, such as the National Oceanic and Atmospheric Administration (NOAA) wind profiler demonstration network (WPDN) (Lee et al. [1995]; Mace et al. [1995]; Mace and Ackerman [1996]). For example, Mace et al. [1995] have shown how to use a kinematic method with a network of wind profilers.

Most of the clear-air radar techniques outlined above can also be applied to other ground-based remote sensing instruments for vertical velocity measurements. However, these instruments do not rely on the backscatter echoes of refractive index irregularities and are not under consideration here. For example, coherent Doppler lidars use a Zenith technique but backscattered echoes are from atmospheric aerosol particles (e.g. Frehlich et al. [1994]). As a second example, Doppler weather radars uses a VAD technique and their targets are mainly hydrometeors (e.g. Doviak and

Zrnič [1993]).

2.3 Vertical Velocity Characteristics

In the following sections we outline some theoretical and observed characteristics of the vertical velocity under different conditions.

2.3.1 Vertical velocity characteristics from theory

Typically, vertical velocity is constrained to be zero at the ground and is assumed to be zero at the tropopause (Bluestein [1992], p. 296). The former is required by the kinematic lower boundary condition: the latter is a dynamic requirement owing to to the reversal of the quasi-horizontal temperature gradient. In spite of this assumption, the tropopause itself does undulate up and down. If w is nonzero, its sign is often the same at all levels in a column of the troposphere. The levels where $\partial w/\partial z$ reverses are called the levels of non-divergence (LND). Although one LND is most often expected, no physical principle is violated if there are more than one. In addition, rising motions ($w > 0$) above a LND must be accompanied by convergence below and compensating divergence aloft. Similarly, sinking motions ($w < 0$) must be accompanied by divergence below and convergence aloft (continuity equation in a non-compressible fluid).

Over a sloping terrain, w may not be zero due to "orographic forcing". This forcing depends on both the horizontal wind speed and the steepness of the slope of the surface. This effect may be so large that orographic effects may completely mask the effects of dynamical forcing.

Many temperature inversions in the lower atmosphere are attributable to sinking motion. When air subsides ($w < 0$), its temperature increases at the dry adiabatic lapse rate $\approx 10°$/km. In some instances, the subsidence continues to a particular level and then the air diverges horizontally. The more rapid the decrease of temperature with height, the more unstable the air, and the greater the vertical

motions.

Based upon basic physics principles, when a volume of air is warmer than its surroundings, the upward pressure-gradient force exceeds the downward gravity force. This results in a positive buoyancy and the air rises: conversely when a volume of air is colder than its surroundings it is negatively buoyant and it sinks (Battan [1979]).

The largest theoretically-determined vertical velocities averaged over a season are on the order of a few millimeters per second. These values occur in the equatorial zone and are associated with the ascending branch of the Hadley cell. Vertical velocity magnitudes decrease considerably in the lower stratosphere. At greater altitudes, the mean vertical motions driven by wave forcing are expected on theoretical grounds to be considerably larger, but nevertheless are $\approx$ 1000 time less than characteristic horizontal mean motions (Fritts and Luo [1995]).

The conceptual model of Mattocks and Bleck [1986] predicts the existence of a transverse circulation around a jet streak. A jet streak is an isotach (line of constant wind speed) maximum embedded in a jet, that is an "intense" (> 15 m s^{-1}), "narrow" (width is at least a half to an order of magnitude less than its length) quasi horizontal or horizontal current of wind that is associated with strong vertical shear (Bluestein [1992]). According to this model, vertical velocities should show ascent above the jet and descent below if measured in the entrance region on the cold side of the jet streak or on the warm side in the exit region. An oppositely directed vertical circulation would be expected for observations made in the warm (entrance) and cold (exit) quadrants.

2.3.2 Some observed vertical velocity characteristics

2.3.2.1 Mean vertical motions

Excepting convective conditions, mean vertical velocities are usually much less than horizontal velocities in the free atmosphere. Large-scale vertical motions are usually less than few centimeters per second. Such small motions are traditionally inferred from synoptic analysis of the large-scale wind field (Panofsky [1951]; Dunn [1991]). Belt and Fuelberg [1982] found that vertical velocities calculated with sub-synoptic scale data were as much as 6 cm s^{-1} larger than those calculated with synoptic scale data in the same conditions.

Assuming that long-term mean vertical velocities ($\bar{w}$) at a single station are representative of large-scale vertical motions, Nastrom et al. [1985] and Larsen et al. [1988] have compared radar-observed mean vertical velocities to vertical velocities obtained from analyzed wind fields produced by the National Meteorological Center (NMC) and the European Centre for Medium-Range Weather Forecast (ECMWF), respectively. Nastrom et al. [1985] concluded that the single-station radar measurements were representative of the synoptic-scale vertical velocities, particularly when the measurements were not contaminated too severely by lee waves. This meteorological "noise" can swamp the synoptic-scale signal at times. Although the observed velocities significantly overestimated the calculated velocities, the signs of the two velocities were generally the same. However, the magnitudes of the w measured over a flat terrain (Flatland, Illinois), where the effects of lee waves are not expected, show large discrepancies with predicted values (Nastrom and VanZandt [1994]; Nastrom et al. [1994]). Instead of being very small, the long-term mean vertical velocity ($\bar{w}$) at Flatland is observed to be downward throughout the troposphere, with values > 5 cm s^{-1} when averaged over periods ranging from hours to months, rather than nearly zero as expected. In addition, $\bar{w}$ is usually observed to be slightly upward in the lower stratosphere.

Similar descending patterns in the troposphere and ascending in the lower stratosphere are sometimes seen by radars at other mid-latitude sites. These results have been reported using the MU radar in Japan (Fukao et al. [1991]) and the SOUSY radar in Germany (Yoe et al. [1992]). Both the MU and the SOUSY radar results have been associated to the proximity to the jet stream, where the $\bar{w}$ reversal occurs at a height near the jet stream maximum. However, when the position of the jet is shifted, that is not usually the case, the pattern reverses, i.e., ascent in the troposphere and descent in the lower stratosphere. It is worth pointing out, that the maximum vertical velocities obtained with the SOUSY radar were a factor of 3 greater than those measured with the MU radar.

The results at tropical sites (Pohnpei [Western Pacific], Christmas Island [Central Pacific] and Piura [Eastern Pacific]), show also similar patterns to those most often seen at mid-latitudes (Balsley et al. [1988]; Gage et al. [1992]; Huaman and Balsley [1996]), except that the stratospheric results at Pohnpei show downward velocities under clear conditions. However, the magnitudes of the tropical mean vertical velocities are very much smaller than those observed at mid-latitudes. "Adiabatic warming associated with the tropospheric subsidence appear to be in approximate balance with calculated radiative cooling to space in the clear atmosphere, but ascending motions appears to be somewhat larger than is needed to balance diabatic heating in the lower stratosphere, assuming a clear atmosphere free of ice or other aerosols" (Gage et al. [1991a]).

2.3.2.2 The variability of vertical velocity

Non-averaged vertical velocity observations in the atmosphere are dominated by short-period internal gravity waves with typical rms amplitudes (σ_w) of 10 to 20 cm s^{-1} (Gage [1990]). Non-averaged w measurements are much more easily realized than measurement of the much smaller long-term mean vertical motions that

requires careful averaging of large numbers of individual measurements (e.g. Gage et al. [1991a]; Balsley et al. [1988])

Motions due to atmospheric turbulence and waves are usually the major source of variance of the measured non-averaged vertical velocities. Observed vertical velocities due to various wave and turbulence motions, while generally smaller than associated horizontal velocities, are far larger (more than 2 orders of magnitude) than the mean vertical motions. Although clear-air radars appear to be suitable for these measurements, the effect of this variability has to be considered in some measurements of synoptic (i.e. longer term) scale vertical velocities (e.g. Nastrom et al. [1985]).

Vertical observations presented by Ecklund et al. [1985] showed, mainly, two conditions: quiet ($\sigma_w < 20$ cm s^{-1}) and active ($\sigma_w > 20$ cm s^{-1}). The variations observed during the active periods, appear to be related to orographic waves that develop as wind flows over rough terrain. Presumably, tilting isentropic surfaces associated with lee waves are responsible for the increased variance (Gage and Nastrom [1990]). The vertical velocity spectra under active conditions are distinctly different from the flat spectra observed under quiet conditions. Specifically, the vertical velocity spectrum under active conditions contains much more low-frequency energy.

Comparisons of w obtained over rough terrain (Platteville, Colorado) and those obtained over a flat terrain (Flatland, Illinois) indicate that terrain indeed does dominate many of the features of mesoscale variability in w. Moreover, the results over a flat terrain do not show the enhanced vertical velocity spectrum (VanZandt et al. [1991]). Also, vertical velocities measured at mid-latitudes (Nastrom and VanZandt [1994]) show much bigger daily variance than those obtained at equatorial sites (Huaman and Balsley [1996]).

An overall view of the variability of vertical velocity fields over Poker Flat,

Alaska, is contained in Balsley and Garello [1985]. Briefly, the short-period vertical velocity variance profiles for the winter months show little change with altitude below 15 km. Above 15 km the profile increases steadily. During the summer months, vertical variance below 12 km is enhanced above the winter values and above 12 km the summer values are lower. The low level increase during the summer months could be due to convection. The decrease at higher altitudes could be due to a wave-breaking process.

2.4 Current concerns with the Zenith technique

As we mentioned earlier, the Zenith technique is currently employed by most clear-air radars. For example, the vertical wind has been measured across the tropical Pacific at all Tropical Pacific Profiler Network (TPPN) sites from the inception of the TPPN network (Balsley et al. [1988]; Gage et al. [1991b, 1991a]; Huaman and Balsley [1996]). In the absence of appreciable cloud cover, typical mean vertical velocities show a slight downward subsidence (≈ -1 cm s^{-1}) throughout the troposphere. While this value is small and difficult to measure, it is, as discussed above, consistent with atmospheric cooling by radiation to space under clear-air conditions. Velocity values under more convective conditions depart greatly from these subsidence values, particularly in the region of so-called "hot towers" associated with deep tropical convection, where upward vertical motions can easily exceed 15 m s^{-1} (i.e. values well over three orders of magnitude larger than clear-air subsidence).

In spite of these reasonable results, there exist a series of puzzling mean vertical velocity measurements (e.g. Balsley and Riddle [1984]; Fukao et al. [1991]; Nastrom and VanZandt [1994]) that bring into question the accuracy of these direct (Zenith technique) radar measurements. For example, Fukao et al. [1991] found that in most of the cases they studied, the direction of the vertical circulation near a Jet stream could not be explained by the slope of the time-averaged potential

temperature surfaces alone, since the combination of the horizontal winds and the slopes of the isentropic surfaces would have led to a prediction of a circulation directly opposed to that observed.

In contrast to these puzzling observations, many case-studies have shown that "instantaneous" (< 5 min averaged) radar vertical motions are consistent with the associated meteorological mesoscale events, such as the cases presented by Low [1996]; May and Rajopadhyaya [1996]; Ralph et al. [1993]; Nastrom et al. [1990a]. These case studies have provided convincing evidence for the applicability of vertical winds, particularly of the "instantaneous" and mesoscale studies. For example, Ralph et al. [1993] observed consistent vertical motions and pressure perturbations associated with a ducted gravity wave.

Ascertaining the validity of vertical velocity measurements is extremely difficult. This difficulty arises because of the lack of a definitive w to compare with. The standard meteorological techniques for estimating the vertical motions (Section 2.2.1) involve derivations from large-scale synoptic analysis, so that such techniques obviously return a broad-scale temporal and spatial average. Comparisons between these techniques and the Zenith radar technique have been made, and show reasonable agreement (Nastrom [1984]; Nastrom and VanZandt [1994]), although there can be contamination from lee-wave effects near mountains (Nastrom et al. [1985]; Larsen et al. [1988]) or from heavy rain (Balsley et al. [1988]). Moreover, geophysical variability places a limit of about 2 cm s^{-1} on clear-air mean vertical motions, averaged over an hour or more (Nastrom et al. [1990b]).

The most often considered factors that could contaminate vertical velocities obtained by Zenith techniques, are outlined in the following paragraphs.

2.4.1 Antenna beam pointing inaccuracies

Only a slight off-vertical tilt of the beam (less than 0.1°) will produce a major contamination of the $\approx$ 1 cm s^{-1} measured vertical wind component by the $\approx$ 10 m s^{-1} horizontal wind. In a first attempt to examine the effects of antenna pointing problems in vertical wind measurements, Huaman and Balsley [1996] made a series of measurements using the Piura (Perú) profiler. This site is the eastern terminus of the TPPN. They found that initial measurements of the vertical wind at Piura had been made using a slightly off-vertical antenna beam. The calculated angle between the true vertical and the antenna pointing direction was extremely small (0.057°). In spite of this, they were able to demonstrate that the contribution of the horizontal wind using this slightly-tilted (2° half-power beam-width) antenna was sufficient to produce almost a factor of two error in the mean vertical wind measurements. The antenna was modified to correct this problem, and the resulting mean vertical wind profiles compared very favorably with comparable profiles at other TPPN sites. Huaman and Balsley [1996] concluded that: (a) such small pointing errors can contaminate long-term mean vertical wind data, and (b) VHF radars are indeed capable of measuring the mean vertical wind with the requisite accuracy in the tropics, provided that one is very careful with the antenna phasing.

2.4.2 Tilted refractive index structures

In general, it can be assumed that anisotropic refractive index structures, are not horizontally stratified and are tilted at small angles (e.g. Röttger [1984]; Röttger et al. [1990b]). This tilt may be the result of the presence of waves (e.g. Gage et al. [1981]; Muschinski [1996]) as well as due to baroclinic disturbances. Although these measured tilt angles are typically small ($\approx 1° - 2°$), the difference between the true vertical velocity and the radial velocity in a nominally vertical beam can be quite significant, as shown by Larsen and Röttger [1991] and Palmer et al.

[1991]. The effects of tilted layers are expected to be proportional to the antenna beam-width (e.g. Muschinski [1996]), i.e., the smaller the beam-width the smaller the contamination on the directly-measured vertical velocity. However, if the tilt angles can be measured, a "corrected" vertical velocity could be obtained (Röttger and Ierkic [1985]). Higher-frequency radars do not suffer from aspect sensitivity effects and, therefore, may avoid some of the problem associated with VHF vertical velocity measurements.

2.4.3 Gravity waves

Nastrom and VanZandt [1994] suggested that the dominant effect causing "apparent" unreasonably large mean vertical motions at mid-latitudes is the effect of small changes in the refractive index induced by gravity waves with a vertical component of propagation. In waves with upward energy propagation (downward phase propagation), the vertical velocity and static stability are negatively correlated: thus, regions of downward-moving air have higher reflectivity on a statistical basis in such waves and will produce apparent downward velocities. This hypothesis is independent of the antenna beam-width. The Nastrom and VanZandt studies incorporated two mid-latitude profilers: (a) the Flatland profiler located near Urbana, Illinois, and (b) a second profiler temporarily located at Liberal, Kansas. They reported much larger subsidence values than those observed at equatorial sites (Balsley et al. [1988]; Gage et al. [1991a]). Their hypothesis has been supported by Hoppe and Fritts [1995] using data from echoes obtained near the polar summer mesopause. The effects of gravity waves on vertical velocities would be also seen by other radar methods (SA, VAD, etc).

2.4.4 Wind field gradients

Recently Palmer et al. [1997a] and Larsen and Palmer [1997], have shown how gradients in the wind field could affect vertical velocity and angle-of-arrival

(AOA) measurements. The effects of divergence on vertical velocity measurements using the Zenith technique, are a function of the beam-width, divergence value, and altitude. These biases are proportional to antenna beam-width, i.e., the smaller the beam-width, the smaller the bias. Divergence will also affect the AOA measurements. Nonetheless, "the bias in the vertical beam radial velocity measurements produced by divergence has the same form and the same effect as the bias that would be expected for aspect-sensitive scatter and inclined refractivity layers" (Larsen and Palmer [1997]). Larsen and Palmer have also shown that vorticity will affect the AOAs but will not bias the vertical velocities. However, if these affected-by-vorticity AOAs are utilized to correct an "unbiased" vertical velocity using Equation 2.4, the corrected vertical motion will be biased by these "fictitious" AOAs.

CHAPTER 3

FOCUSING IN ON THE VERTICAL VELOCITY PROBLEM

Given the significant discrepancies in vertical velocity measurements by different radar techniques, particularly in the long-term mean values, we have implemented a variety of techniques for measuring atmospheric vertical motions. All of these measurements have been obtained concurrently using the Jicamarca VHF radar. Our goals are to intercompare the different techniques and to study the validity of various explanations that attempt to account for vertical velocity discrepancies (Chapter 6).

Two Zenith techniques have been implemented for these studies. One technique uses a very narrow antenna beam-width, while the other uses a wide antenna beam-width. Both beams are nearly collocated (the distance between the two antennas is about 65 m from center-to-center). Since both techniques make use of orthogonal polarizations, their measurements can be considered independent. Using these two antenna beam-widths, we study two possible causes of vertical velocity discrepancies that depend on the antenna beam-width, i.e., tilted refractive index structures and wind field divergence.

We have also implemented a series of SA techniques. Most of these techniques (the so-called SA "classical" techniques) correct estimates of vertical velocity that have been contaminated by horizontal velocity "leakage". This leakage is associated with both the antenna pointing inaccuracies as well as with the existence of tilted refractive index structures. Note that other type of SA techniques infer vertical velocities by using a "quasi" DBS approach in a multi-beam configuration. The usage of many beams allows the averaging out of effects that depend on beam-pointing

direction.

Gravity wave effects are not directly tested in the measurements. This is mainly because wave dynamics over Jicamarca (near the Equator) are expected to be much different than conditions present at mid-latitudes. For example, the hourly mean variance at the Piura ST radar ($\approx$ 1000 km North of Jicamarca) is $\approx$ 20 times less than in mid-latitudes (Huaman and Balsley [1996]). Using Nastrom and VanZandt [1994] expressions for the effects of gravity waves, vertical velocity biases near the Equator should be a fraction of a centimeter per second.

Since the corrected vertical velocity estimates obtained by the classical SA methods require the measurement of horizontal velocities and AOAs (see Equation 2.4), we first study the accuracy of these two measurements. Studies of horizontal velocities (Chapter 4), given the large number of SA techniques found in the literature, help us to determine which horizontal velocity techniques are most appropriate to vertical velocity corrections.

During these horizontal velocity comparisons, in view of the lack of an absolute standard, we face the problem of using a reference profile. A good choice would have been to use the widely-accepted DBS measurements. Unfortunately, DBS measurements were not obtained as part of these studies. This omission was because of the relatively small maximum angles of the oblique beams ($< 6°$) at Jicamarca. Tsuda et al. [1986] have shown that for angles $< 10°$ under anisotropic scattering conditions, the DBS horizontal velocity will be underestimated because the effective beam direction will be closer to zenith than the real beam direction. This underestimate is caused by the anisotropic angular pattern of the scattering/reflecting inhomogeneities (aspect sensitivity). Therefore, we have arbitrarily chosen one of the most commonly used full correlation analysis technique as a reference.

The AOA studies that will be discussed in Chapter 5 will help us to: (1)

correctly interpret the AOA measurements obtained with SA systems, and (2) properly use the measurements in vertical velocity corrections. For example, we will study the geometrical effects suggested by Vincent et al. [1987] and May [1993] on AOAs measured with "non-collocated" (or bistatic) SA systems.

CHAPTER 4

A STATISTICAL COMPARISON OF HORIZONTAL WINDS OBTAINED BY A VARIETY OF SPACED ANTENNA TECHNIQUES USING THE JICAMARCA VHF RADAR

4.1 Introduction

Horizontal wind velocity profiles can be measured in the troposphere and lower stratosphere using very high frequency (VHF) radars. There are two main methods that can be used in these measurements. One of these methods, the Doppler or Doppler Beam Swinging (DBS) method (Balsley [1981]), uses three or more narrow beams pointed in different directions. In this method, the wind velocity vector is determined by measuring the Doppler shift of the signal arising from the radial motion of the refractive index variations in the sampled scattering volume. The DBS method is used by most operational wind profilers (e.g. Strauch et al. [1984]; Gage et al. [1991b]). The second method, called the spaced-antenna (SA) method, was originally developed to look at ionospheric returns (Briggs et al. [1950]) and uses a vertically-directed antenna for transmission and three or more horizontally spaced receiving antennas (Larsen and Röttger [1989]; Hocking et al. [1989]). The SA method makes use of the fact that backscattered signal is the diffraction pattern of the illuminated refractive index irregularities. This diffraction pattern moves across the ground at twice the velocity of the scatterers. Normally, winds are calculated from the signals detected by each of the three (or more) antennas using correlation techniques (e.g. Briggs [1984]; Meek [1980]). Since the diffracted signal is sampled at more than a single location by this method, it should not be surprising that the amount of information that can be extracted increases. Röttger and Larsen [1990]

give a detailed review of the DBS and SA methods.

Practically speaking, the DBS analysis can only be analyzed by the estimation of the moments of the radial velocity estimates, although the performance of various moment estimators may vary (May et al. [1989]). On the other hand, there is a number of techniques that can be applied to the SA returns. Some of these techniques are variations of the classical approach presented by Briggs [1984]. Other techniques are more closely related to the DBS method in the sense that they obtain the velocity vector from "radial velocities". Examples of such techniques include post beam steering (Röttger and Ierkic [1985]), post statistics steering (Kudeki and Woodman [1990]), and time domain interferometry (Vandepeer and Reid [1995]).

In general, the techniques differ not only in how the signals are interpreted, but also in which domain they are processed (i.e. time or frequency). Frequency-domain techniques offer several advantages over their time-domain counterparts. For example, frequency-domain techniques allow relatively easy removal of undesired frequency contamination such at that resulting from sea clutter, system interference and/or precipitation echoes. Corrections for ground clutter and noise are also simpler in the frequency domain. Moreover, spectral analysis is also computationally more efficient than correlation analysis.

Comparisons of a few SA techniques has been given by Franke et al. [1990], Sürücü et al. [1992] and Holdsworth [1997] using either MF radar results or computer models. Additional comparisons with the VHF MU radar (see Fukao et al. [1990]) have been made by Van Baelen et al. [1991], Sheppard et al. [1993] and Brown et al. [1995].

We present here a comparison of a variety of VHF SA techniques using a total of 13 days worth of data obtained using the multi-module antenna at the Jicamarca Radar Observatory in Perú. These comparisons were made in the troposphere and lower stratosphere and concern only horizontal wind measurements. We have

attempted to quantify the differences of the many SA techniques and to demonstrate some of the specific advantages and disadvantages between them.

A brief introduction to the relevant SA techniques is presented in Section 4.2, along with details of the implementation procedures. A description of the antenna setup and the experimental parameters appears in Section 4.3. Results from two separate observational periods (i.e. "quiet" and "active" wind conditions) are presented in Section 4.4. Explanation of possible causes of differences determined between the techniques is discussed in Section 4.5, and a summary of the results is presented in Section 4.6.

4.2 Spaced-Antenna Techniques

In the literature, one can find a variety of multi-receiver techniques whose differences depend either on how they are implemented (time or frequency domain) or on how the frequency information is used. These techniques have been given different names, e.g., full correlation analysis (FCA), full spectral analysis (FSA), radar interferometry (RI), spatial interferometry (SI), frequency domain interferometry (FDI), time-domain interferometry (TDI), imaging Doppler interferometry (IDI), etc.

To avoid confusion in presenting our results, we will describe briefly only those techniques that use a single-frequency, multiple-receiving antennas, and that are a part of this study.

To infer horizontal winds, the majority of the SA systems calculate the motion of the diffraction pattern across the ground associated with the motions of the atmospheric scatterers in the resolution volume. If the scatterers are assumed frozen, the diffraction pattern is advected with a velocity equal to twice the horizontal wind in that volume ("point source effect", e.g. Briggs [1980]). By observing the maximum value of the cross correlation of signals from several receiver pairs, the wind

can be estimated using the time lag of that maximum. However, the decorrelation of the received signals caused by the relative motion of the scatterers, i.e., those motions arising from turbulent eddies, causes an error that leads to an overestimation of the wind speed. It is at this point that we encounter our first velocity estimator that will result in a so-called "apparent" velocity. Depending on how this decorrelation is treated, the domain used (time or frequency), and/or the assumptions made, one can find different techniques in the literature. We have implemented some of these techniques and classified them by either time or frequency domain techniques

4.2.1 Time-Domain techniques

Time domain techniques dealing with the decorrelation of signals follow essentially a full correlation analysis (FCA). FCA allows not only for the temporal but also for the spatial decorrelation of the diffraction pattern. FCA uses the complex cross-correlation functions (CCF) from all the receiver combinations, and also includes the autocorrelation function (ACF). CCFs are defined in the spatial and temporal domain by:

$$\rho(\varepsilon, \eta, \tau) = \frac{< v^*(x, y, t) \cdot v(x + \varepsilon, y + \eta, t - \tau) >}{< v^2(x, y, t) >} \tag{4.1}$$

where $v(x, y, t)$ is the voltage measured at time t and at antenna located at (x, y), and <> denotes ensemble averaging. The spatial displacements and time delay are denoted by ε, η, and τ.

Depending upon which functional form of the diffraction patterns is assumed, the following techniques have been described in the literature:

• FCA by Briggs (here: FCAB). Briggs [1984] gives a detailed formulation and a "recipe" to implement this technique. The basic assumption in this case is that contours of equal spatial and temporal correlation of the ground diffraction pattern can be approximated by a family of ellipsoids of the general form:

$$|\rho(\varepsilon, \eta, \tau)| = \rho(A\varepsilon^2 + B\eta^2 + C\tau^2 + 2F\varepsilon\tau + 2G\eta\tau + 2H\varepsilon\eta) \tag{4.2}$$

and, using intersections of the ACF and CCFs, the A, B, C, H, F, G parameters are obtained. A a "true" horizontal velocity can be calculated using these parameters.

- FCA by Meek (here: FCAM). This technique has been described by Meek [1980]. The concepts are similar to FCAB, but with a Gaussian correlation function assumed for the magnitude of the CCFs. Only the width of the mean ACF and the positions and magnitudes of the maxima in the CCF are required to calculate a "true" velocity (here: FCAMt). We have implemented this technique by fitting a Gaussian function to the magnitudes of the ACF and CCFs.
- FCA by Kudeki (here: FCAK). Kudeki [1995] derived the following explicit formula

$$|\rho(\varepsilon,\eta,\tau)| = 1 - A\varepsilon^2 - B\eta^2 - C\tau^2 - 2F\varepsilon\tau - 2G\eta\tau - 2H\varepsilon\eta \tag{4.3}$$

which is valid for "small" values of ε, η and τ. Then, using measured CCF samples at, say, $(\varepsilon,\eta,\tau) = (\varepsilon_0,0,0)$, $(0,\eta_0,0)$, ..., $(\varepsilon_0,\eta_0,\tau_0)$, the A, B, C, F, G, H parameters can be obtained using a single value decomposition (SVD) solution to the overdetermined set of equations. Knowing these parameters, a "true" velocity can be calculated in a way similar to the FCAB "true" velocity.

- Relaxing the assumption of the elliptical symmetry in the shape of the CCFs, Woodman [1995] arrived at a new approach (here: FCAW) to determine a "true" velocity for statistically isotropic patterns, as well as for anisotropic patterns when the receiving antennas are aligned along the axis of symmetry. For any baseline $\mathbf{d} = (\varepsilon,\eta)$ the intersection of the ACF, $|\rho(0,0,\tau)|$, with the CCF, $|\rho(\varepsilon,\eta,\tau)|$ occurs at a delay, τ_c, and the projection of the wind velocity, $\mathbf{v}$ over $\mathbf{d}$ is given by,

$$v_d = \frac{\mathbf{v}.\mathbf{d}}{2\,d} = \frac{d}{4\tau_c} \tag{4.4}$$

by evaluating the projection of two or more different directions, it is possible to determine the two-dimensional wind vector. Note that Armstrong and Coles [1972] and Briggs and Maude [1978] also used the "crossings" of CCFs to estimate drift velocities.

• Following the work by Doviak et al. [1996], Holloway et al. [1997a] has presented what are essentially two separate time-domain techniques for estimating the wind velocities for horizontally isotropic scattering (i.e., $A = B$ and $H = 0$ in Equation 4.2). One of these techniques uses the crossings of the ACF and the CCFs to determine the wind velocity using Equation 4.4. The other technique uses the ratio of the CCFs to the ACF, $|\rho(\varepsilon, \eta, \tau)|/|\rho(0, 0, \tau)|$, to calculate the wind velocity. Taking the log of this ratio, the slope S of this curve (with respect to time) is given by:

$$S = \beta_h^2 \, d \, v_d \tag{4.5}$$

and its intercept b at the $\tau = 0$ axis is given by:

$$b = -\frac{\beta_h^2 \, d^2}{4} \tag{4.6}$$

where β_h is related to the scale length of the diffraction pattern [Holloway et al., 1997a, Equation 5].

We have implemented two versions of the latter of these two techniques. The first version uses the "raw" ACF and CCFs (here: FCAH(r)) and the projected velocity is obtained using Equation 4.5 and Equation 4.6 and is given by:

$$v_d = -\frac{d \, S}{4 \, b} \tag{4.7}$$

In the second version, Holloway et al. [1997a] fit a Gaussian function to the magnitudes of the ACF and CCFs (here: FCAH(p)), and the projected velocity is given by:

$$v_d = \frac{d\hat{\tau}_p}{2\hat{\tau}_p^2 - 4 \ln (A_d/A_0)\hat{\tau}_c^2} \tag{4.8}$$

where $\hat{\tau}_p$ is the estimated delay to the peak, $\hat{\tau}_c$ is the estimated width, A_d is the magnitude of the Gaussian fitted to the CCF used, and A_0 is the magnitude of the ACF (≈ 1). Once a Gaussian function is assumed, Equation 4.4 and Equation 4.7 are the same and equal to Equation 4.8.

The last time-domain technique that we have implemented is more closely related to the DBS method than to a classical SA (FCA) approach. This technique is called Time-Domain Interferometry (here: TDI) introduced first by Vandepeer and Reid [1995] and modeled later by Thorsen et al. [1997]. This technique is based on the angle-of-arrival (AOA) and Doppler velocity (v_r) measurements from the zeroth and first lags of the CCFs and ACF functions. Note that both the time-domain techniques described before, and the TDI technique can utilize complementary data sets; the former techniques use only the magnitude of the CCFs, while the latter use only the phases. Basically the three-dimensional TDI wind velocity is obtained from a set of equations of the form,

$$v_{r_t} = -(l_t\, u + m_t\, v + (1 - l_t^2 - m_t^2)^{1/2}\, w) \tag{4.9}$$

where u, v and w are the zonal, meridional and vertical components of the wind velocity; v_{r_t} is the radial velocity and l_t and m_t the zonal and meridional direction cosines at some time t. The wind field is then determined from a linear least squares fit to a system of equations similar to Equation 4.9. Since the phases of CCFs are used, a calibration procedure is necessary in order to remove systematic biases in the receiving lines.

4.2.2 Frequency-Domain techniques

Since cross-correlation and cross-spectrum functions (CSFs) of complex receiver signals are Fourier transform pairs, it is expected that there should be frequency-domain estimators analogous to the time-domain techniques.

The first frequency-domain technique we have implemented is the "Radar Interferometry" technique (here: RIV) introduced by Van Baelen and Richmond [1991], where the horizontal wind field is determined from the phase slope of CSFs. By the "shift theorem" of Fourier transforms, one can view the RIV horizontal velocity as the Fourier equivalent of the "apparent" velocity described above, i.e.

the velocity obtained when turbulence is neglected.

The second frequency-domain technique we have implemented is the full spectral analysis (FSA) technique (Briggs and Vincent [1992]; Sheppard and Larsen [1992]). Here, we have used the FSA technique developed by Briggs and Vincent [1992], where they assumed a Gaussian form for the CCF defined in Equation 4.2, and Fourier transformed it to get a CSF. In order to derive the *A*, *B*, *C*, *F*, *G*, *H* parameters, determinations of the phase slope, the amplitude, and the width of the CSFs are needed. The FSA analysis, like its time-domain analogous FCA, produces two velocities estimates, namely, an "apparent" and a "true" velocity. The "apparent" velocity is the same as the velocity given by the RIV technique. In the present study, when we mention the "FSA velocity", we are referring to the "true" estimate.

Holloway et al. [1997a] also presents a frequency-domain technique that is valid for horizontally isotropic scattering (here: RIH). The RIH projected velocity is obtained from the expression:

$$v_d = \frac{2\, k_0\, S_\phi\, \sigma_v^2}{\beta_h^2\, d} \tag{4.10}$$

where S_ϕ is the phase slope, σ_v is the width of the cross-spectrum; β_h is a parameter described earlier, and k_0 is the radar wave number. Here, the turbulence contribution is included in σ_v. In the frequency domain, the parameter β_h is obtained by a measurement of S_ϕ, σ_v and the normalized power of the cross spectra (P_d).

$$\beta_h = \frac{1}{d}\left[2\, S_\phi^2\, \sigma_v^2 - 4 \ln P_d\right]^{1/2} \tag{4.11}$$

In the techniques just described (RIV, FSA, RIH), we have fitted a Gaussian function to the magnitude, and a line to the phase slope of the cross-spectra functions. As recommended by Sheppard et al. [1993], we have smoothed the CSF magnitudes in order to prevent excessive fitting failures and to improve the spectral width estimation.

The last frequency-domain technique under study is the post statistics steering (here: PSS) (Kudeki and Woodman [1990]; Palmer et al. [1990]; Sürücü et al. [1992]; Palmer et al. [1993]). PSS like the TDI technique, is more closely related to the DBS method than to a classical SA approach. This technique synthesizes desired beam directions using simple arithmetic operations with measured CSFs. Once the beams are synthesized, the moments can be calculated using a first-moment method (FM) such as the one described by May and Strauch [1989], and the derivation of the wind field is analogous to the DBS method (e.g. Strauch et al. [1984]). The "apparent" beam directions (ignoring aspect sensitivity), are calculated considering the geometry involved, the height sampled and the transmitting and receiving antenna patterns. More details on the implementation of the PSS techniques can be found in Appendix D.

In Table 4.1 we summarize the SA techniques incorporated in our study. We have listed in Table 4.1 technique abbreviations, a reference, the type of horizontal scattering considered, whether turbulence is considered or not, the type of fitting we have performed in the CCFs or CSFs, and the domain. We have calculated the time-domain "apparent" velocities using the parameters from the FCAM technique and have labeled them FCAMa.

4.3 Jicamarca Experiment

We have used the Jicamarca radar for our intercomparison studies. The radar operates at $\approx$ 50 MHz and has an $\approx$ 300 m by 300 m antenna array. The antenna is composed of 64 separate modules. Each module consists of a 12 by 12 array of cross-polarized half-wave dipoles. Figure 4.1 shows the antenna configuration for our measurements. For transmission, we phased four modules (section D) to point precisely vertically, with a one-way half-power beam width (HPBW) of $\approx 4°$. Antennas A, B, C were used for reception. Each receiving antenna was

Table 4.1: SA technique abbreviations and some characteristics.

Technique	Reference	Scattering[a]	Turbulence	Fitting	Domain
FCAB	Briggs [1984]	Anisotropic	Yes	None	Time
FCAMa	Briggs [1984]	Isotropic[b]	No	Gaussian	Time
FCAMt	Meek [1980]	Anisotropic	Yes	Gaussian	Time
FCAK	Kudeki [1995]	Anisotropic	Yes	None	Time
FCAW	Woodman [1995]	Isotropic[b]	Yes	None	Time
FCAH(r)	Holloway et al. [1997a]	Isotropic[b]	Yes	None	Time
FCAH(p)	Holloway et al. [1997a]	Isotropic[b]	Yes	Gaussian	Time
TDI	Vandepeer and Reid [1995]	Anisotropic	Yes	None	Time
RIV	Van Baelen et al. [1991]	Isotropic[b]	No	G/L[c]	Frequency
FSA	Briggs and Vincent [1992]	Anisotropic	Yes	G/L[c]	Frequency
RIH	Holloway et al. [1997a]	Isotropic[b]	Yes	G/L[c]	Frequency
PSS	Kudeki and Woodman [1990]	Isotropic[b]	Yes	None	Frequency

[a]Type of scattering in the horizontal plane
[b]or Anisotropic with an axis aligned with the velocity vector
[c]Gaussian/Linear for amplitude/phase

pointed on-axis ($\approx -1.46°$ from vertical along the the y-axis. i.e. toward the SW) with an $\approx 8.2°$ one-way HPBW. We did not use the entire Jicamarca antenna for transmission, since this would not have been efficient for the SA measurements (e.g. Royrvik [1983]). The outputs from each of the antennas used in our study were independently received, sampled, and recorded. Ground clutter effects were reduced separately in both the real and imaginary signal channels. To do this, we calculated the mean complex voltages in each of the blocks used for the incoherent averaging. We then fitted a second degree polynomial to the real and imaginary mean values. Ground clutter effects were removed by subtracting this function from the voltages to provide time series for subsequent analysis.

We obtained a multi-day data set during two different periods separated by about two months. An 8-day run was taken in January, 97 (from January 20, 09:05 LT to January 27, 10:30 LT) and a 5-day run was taken in March, 1997 (from March 25, 16:50 LT to March 29, 22:15 LT). The operational parameters for each data set are presented in Table 4.2.

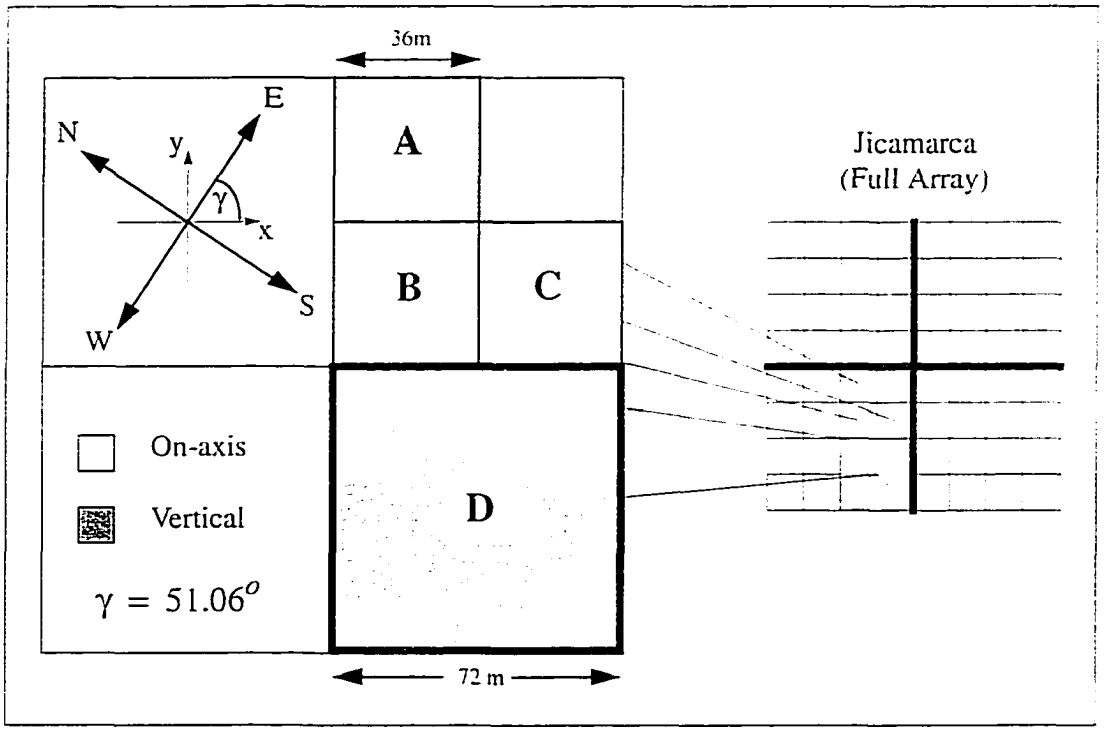


Figure 4.1. Antenna setup for the SA experiment at Jicamarca. The transmitting antenna is the D module, pointing vertically; and the receiving antennas are A,B and C modules. Notice that neither of the antenna axe, (x, y), are aligned with the geographic coordinates (E, N). Moreover, the antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest.

Independent measurements were obtained every ≈ 3 minutes during these periods for all heights between 3 km and 21 km, for each antenna/receiver set. This combined data set was used to analyze all of the techniques studied, excepting the TDI technique. For the TDI analysis, we broke the time series used above into 4 separate sets. Each set consisted of 64/32 points in the quiet/active period (see Table 4.2). Each TDI velocity estimate was obtained solving a set of eight AOA/v_r measurements (see Equation 4.9). Thus, TDI velocity profiles were obtained $\approx$ every 6 minutes.

A calibration procedure was performed before and after each data set to determine the phase differences between the receiving lines. Briefly, a common signal (an output from one of the antennas or amplified sky noise) was fed in parallel to

Table 4.2: Operating parameters for the experiments conducted at Jicamarca

Parameter	Data Set 1	Data set 2
Starting time	20-Jan-97	25-Mar-97
Ending time	27-Jan-97	29-Mar-97
Pulse repetition period	1 ms	2 ms
Number of coherent integrations	128	128
Number of incoherent integrations	6	6
Number of points	256	128
Effective sampling time	256 ms	256 ms
Height resolution	0.45 km	0.45 km
Initial height	3 km	3 km
Number of heights	40	40
Polarization	Down(x)	Up(y)
Transmitting power	100 kW	100 kW
Atmospheric Conditions	quiet	active

each of the receiving lines (cables, front-ends, amplifiers and receivers) and the data was recorded and processed with parameters similar to those in Table 4.2. Phase differences determined from this process were used in the subsequent analysis.

The center-to-center spacing of the receiving antennas were 36 m (AB), 36 m (BC) and 50 m (AC). Such spacings are reasonable for tropospheric and stratospheric studies at VHF (e.g. Larsen and Röttger [1989]).

4.4 Results

The results presented here have been subjected to two "cleaning" processes. The first process, which was essentially a "pre-cleaning", rejected data that did not satisfy the requirements of the technique used (e.g., low correlation of the maxima or intersections of the CCFs in some of the time-domain techniques or poor Gaussian/linear fitting of the magnitude/phase of the CSFs in most of the frequency-domain techniques). Also, a common criterion for all the techniques was to reject estimates with low SNR (< -6 dB). A second "post-cleaning" was performed at the output stages of each technique to eliminate outliers. Data lying outside the range $\bar{x} \pm 2\sigma$ were rejected, where $\bar{x}$ was a 30-minute smooth estimate (including outliers)

and σ^2 was the variance of the data set $x - \bar{x}$.

In order to quantify the agreement of the different SA techniques, we adopted a "reference profile". This profile, which facilitated our intercomparisons, was chosen somewhat arbitrarily. We elected to use the FCAM "true" velocity (here: FCAMt) as our reference. The FCAMt technique had less failure points than the other techniques, following each of the two cleaning stages and exhibited a greater consistency in the measurements. In addition, this technique is one of the most widely used SA techniques (Franke et al. [1990]; Sürücü et al. [1992]; Thorsen et al. [1997]) for wind estimation.

We have divided our comparisons into two periods: a "quiet" period, with low tropospheric wind velocities (≤ 10 m s^{-1}), and an "active" period where winds were much stronger. These comparisons are described separately in the following paragraphs.

4.4.1 Quiet period results

The quiet period corresponds to the 8-day data set taken in January, 1997. Results showing the mean horizontal velocity profiles (magnitude and direction), and relative magnitude differences obtained using twelve different techniques, along with the mean signal-to-noise ratio (SNR) profile) are plotted in Figure 4.2. Each of the separate techniques is indicated by separate symbols, as shown in the figure caption.

Examination of Figure 4.2 shows that the magnitudes of the observed winds are typically < 15 m s^{-1} and increase slightly with increasing altitude. All techniques show a good agreement on the direction of the horizontal wind, with a mean westward wind for the observable regions between $\approx$ 8.5-21 km, and a mean south-westward wind for the region below $\approx$ 8.5 km.

The relative magnitude difference between techniques, Δ, is defined as

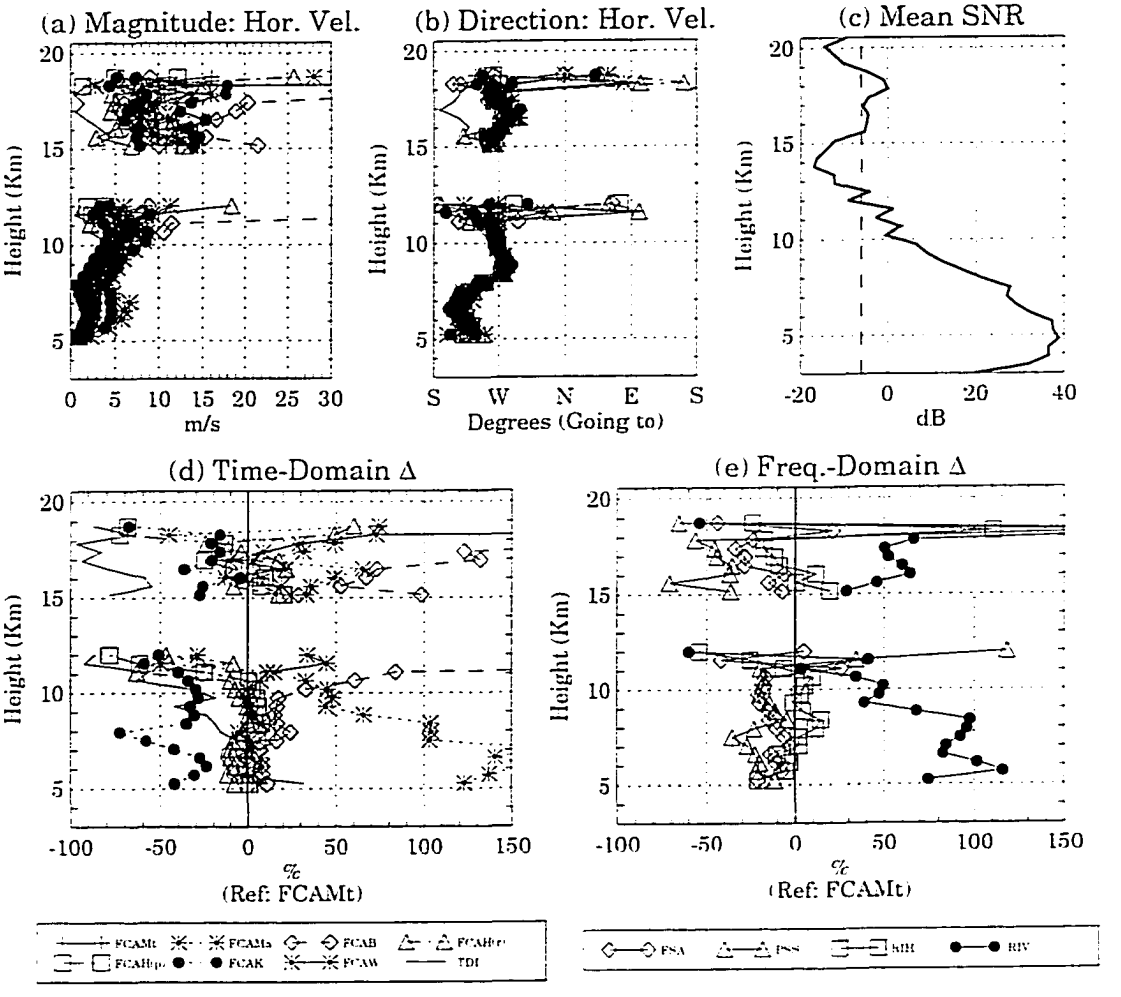


Figure 4.2. Mean horizontal velocity profiles from all techniques for the quiet period (20-Jan-97 to 27-Jan-97). (a) Magnitude, (b) Direction and (c) Mean SNR profile (the vertical dashed line represents the SNR threshold used to reject data). Relative magnitude difference in percent (Δ) for (d) time-domain techniques and (e) frequency-domain techniques at the bottom. See Section 4.2 for description of the different techniques.

$(\sqrt{u_i^2+v_i^2}/\sqrt{u_r^2+v_r^2}-1)$, where (u_i, v_i) and (u_r, v_r) represent respectively the zonal and meridional components of the ith technique and the reference technique r (i.e., FCAMt). This relative difference, shown in Figure 4.2 (d)-(e), shows that the overestimated measurements are clearly coming from the "apparent" velocities (FCAMa, RIV), while the underestimated velocities are, primarily, from some frequency domain techniques dealing with turbulence (FSA, PSS) as well as the FCAK and TDI techniques. The FCAW, FCAH(r) and FCAH(p) appear to correspond most closely to the reference. Note that, for all of the techniques shown, the biases (Δ) appear to increase with decreasing SNR.

We have generated a series of statistical velocity intercomparisons using correlation coefficients (ρ_d), mean differences (μ_d), standard deviation of the differences (σ_d) and percent of failures (α), again using the FCAMt as a reference velocity. These parameters were obtained from scatter plots and histograms similar to the single example presented in Figure 4.3. In this figure, we have compared 3-minute zonal velocity FSA estimates to our reference (FCAMt) for both high SNR heights (5-10 km) and low SNR heights (16-18 km). At high SNR heights (Figure 4.3 (a)-(b)) this comparison shows a high correlation coefficient ($\rho_d \approx 0.85$), a close-to-zero μ_d ($= 0.343$ m s^{-1} and represented by the vertical dashed line in the histogram), and a relatively small σ_d (1.78 m s^{-1}) between these two techniques. However, at low SNR heights (Figure 4.3 (c)-(d)) the comparison gets worse (a lower ρ_d, larger μ_d and larger σ_d). In addition, the FSA measurements show a clear underestimation (Figure 4.3-(c)) and a noisier histogram (Figure 4.3 (d)). The latter result is clearly due to the small number of valid measurements at heights of low SNR.

We have obtained comparable statistical results for all the techniques for both zonal and meridional hourly wind velocities. These results are summarized for the zonal wind velocity during the quiet period for three height ranges in Figure 4.4. Similar results for the meridional wind velocity under quiet conditions are plotted in

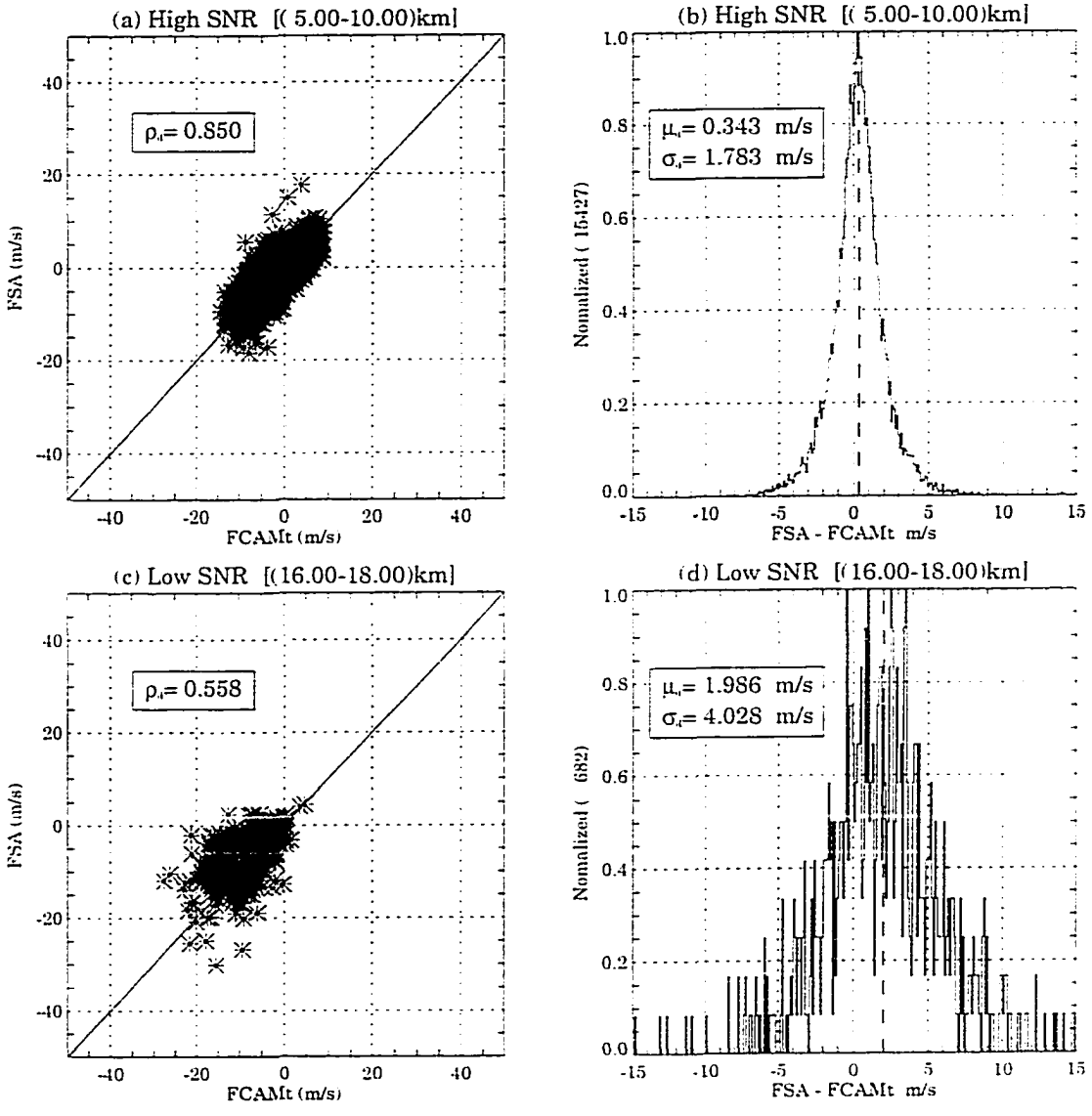


Figure 4.3. Examples of correlation and histogram comparisons of 3-minute zonal velocities obtained by the FCAMt and the FSA SA techniques during the quiet period. (a) and (b) High SNR (5-10 km), and (c) and (d) Low SNR (16-18 km) comparisons. ρ_d represents the correlation coefficient. μ_d the mean and σ_d the standard deviation of the velocity differences (FCAMt - FSA).

Figure 4.5. The techniques being compared to the FCAMt reference are shown in the y-axis of the plots. The three altitude ranges include: (1) all the heights (3-21 km), (2) only tropospheric heights with high SNR (5-10 km), and (3) just near-tropopause heights (16-18 km) where the SNR is low ($\approx$ 0 dB). Similar results for the meridional wind velocity under quiet conditions are plotted in Figure 4.5.

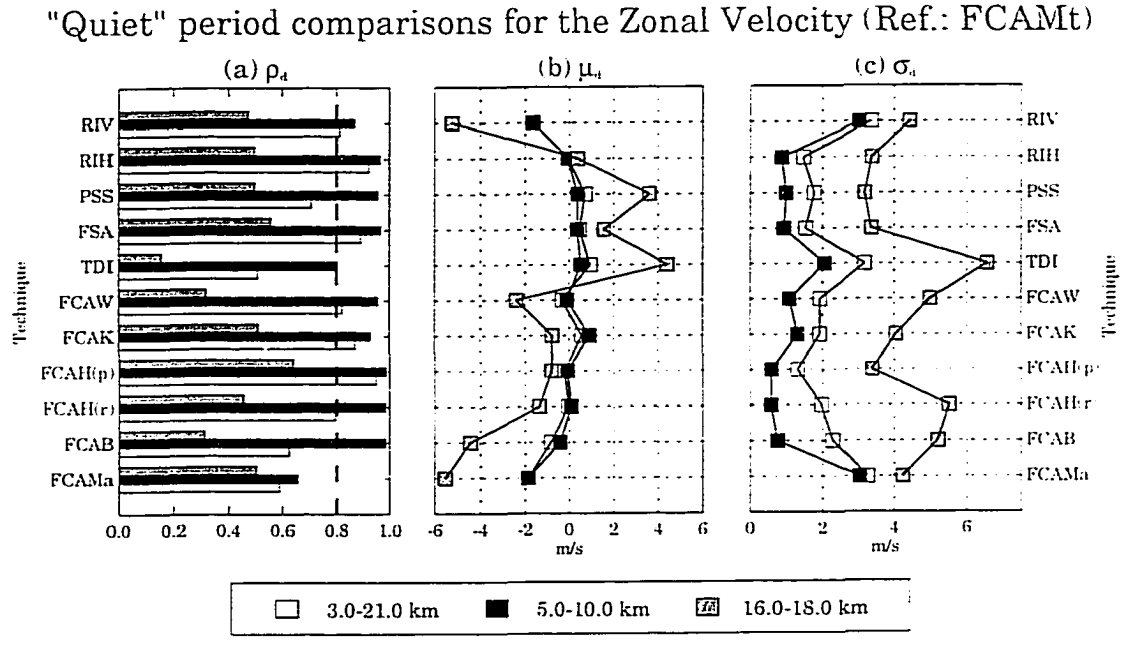


Figure 4.4. Statistical comparisons of hourly-average zonal velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray): during the quiet period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences.

Examination of Figures 4.4 and 4.5 shows reasonable agreement between most of the SA techniques for the 5-10 km quiet period (black squares) with some exceptions, both wind component comparisons show a high correlation ($\rho_d > 0.8$), a small mean difference ($|\mu_d| < \pm$ 1 m s^{-1}) and a small standard deviation of the differences ($\sigma_d <$ 2 m s^{-1}). The agreement gets worse in the 16-18 km range (dark gray squares), where ρ_d decreases ($\leq$ 0.6), μ_d has more variability ($\pm$4 m s^{-1} in the zonal comparisons), and σ_d is larger ($>$ 3 m s^{-1}). This decreased agreement is consistent with the increased statistical errors as SNR decreases. Note that the

"Quiet" period comparisons for the Meridional Velocity (Ref.: FCAMt)

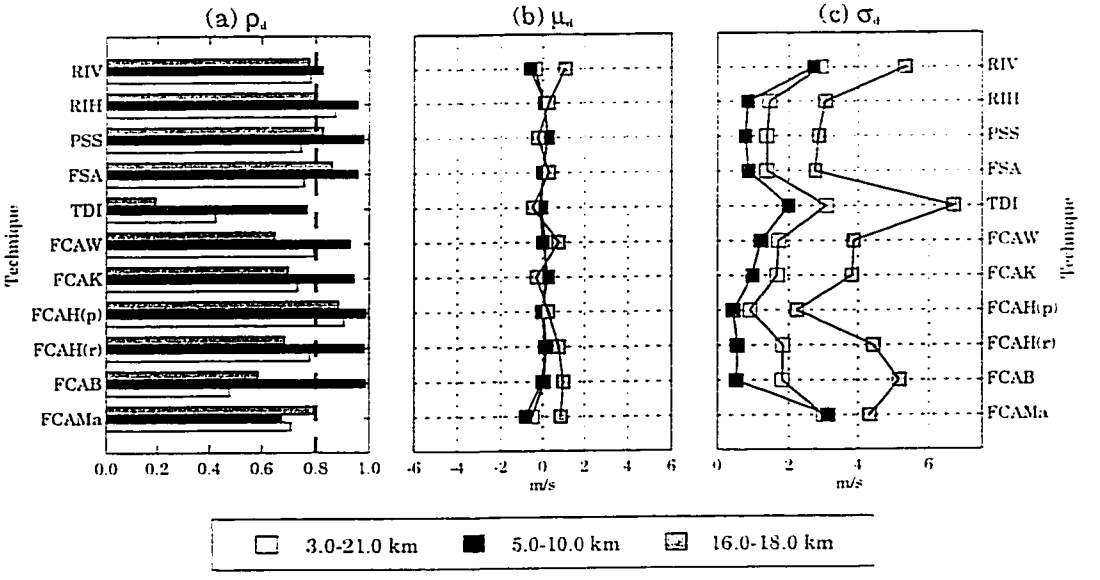


Figure 4.5. Statistical comparisons of hourly-average meridional velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray): during the quiet period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences.

meridional comparisons for this region are better, with smaller μ_d than the zonal comparisons. Note also that the TDI technique shows the poorest agreement with the reference (FCAMt), particularly in σ_d.

4.4.2 Active period results

The active period analysis corresponds to the 5-day data set taken in March, 1997. Figure 4.6 shows similar results to Figure 4.2 but for the active period. The magnitudes of the winds are much stronger than those in Figure 4.2. Again, all the different techniques show a good agreement in the direction of the wind, with a mean westward wind for the $\approx$ 16-18 km region, and a south-eastward mean wind at the lower heights. The relative magnitude differences, Δ, have the same features as the quiet period, except that the FCAK results are closer to the reference, particularly in the 16-18 km height range, and FCAW, FCAH(r) and FCAH(p) results show a slight overestimation that increase with increasing height. Again for all the techniques,

biases increase with decreasing SNR.

Statistical comparisons similar to those presented for the quiet period are presented next. In Figure 4.7 we present examples of scatter plots and histograms, similar to Figure 4.3, for the FCAMt and FSA comparison under the active period. Much larger velocities, somewhat higher values of ρ_d, and larger values of σ_d relative to Figure 4.3 are apparent in this figure. The higher correlation relates to smaller relative errors in velocity even though σ_d has increased. The underestimation by the near-tropopause FSA measurements, is also clear in Figure 4.7 (c). This underestimation is also seen in the positive displacement of the peak in Figure 4.7 (d). Note the skewness in this figure, which probably is due to the different degrees of underestimation at different heights (see Figure 4.6 (e)).

The zonal and meridional hourly results for this active period are shown in Figure 4.8 and Figure 4.9 respectively. Both figures have the same format as Figure 4.4. In the 5-10 km region (black squares) the high correlation persists in most of the techniques, although μ_d and σ_d have increased ($|\mu_d| < 2$ m s^{-1}, $\sigma_d \approx$ 2 m s^{-1}). For the 16-18 km region (dark gray squares), the two wind components have similar ρ_d and σ_d, but μ_d is larger for the zonal component where this zonal component is greater than the meridional wind (see Figure 4.6 (a)-(b)). Again the poor intercomparison appears to lie with the TDI technique.

In Figure 4.10 we present a bar chart of 3-minute zonal velocity "failures" (α, in percent), for the three range sets used previously, during both (quiet and active) periods. The percent of failures here is defined as, the ratio of the measurements rejected over the total number of measurements, by 100. The α values represent the data rejected in the two cleaning stages mentioned earlier, i.e., in "pre-cleaning" and "post-cleaning". Figure 4.10 (a) shows the failures for the quiet period, while Figure 4.10 (b) shows failures for the active period. The lowest percent of failures occurs in the FCAMt measurements. The percent of failures in the 5-10 km region, for all

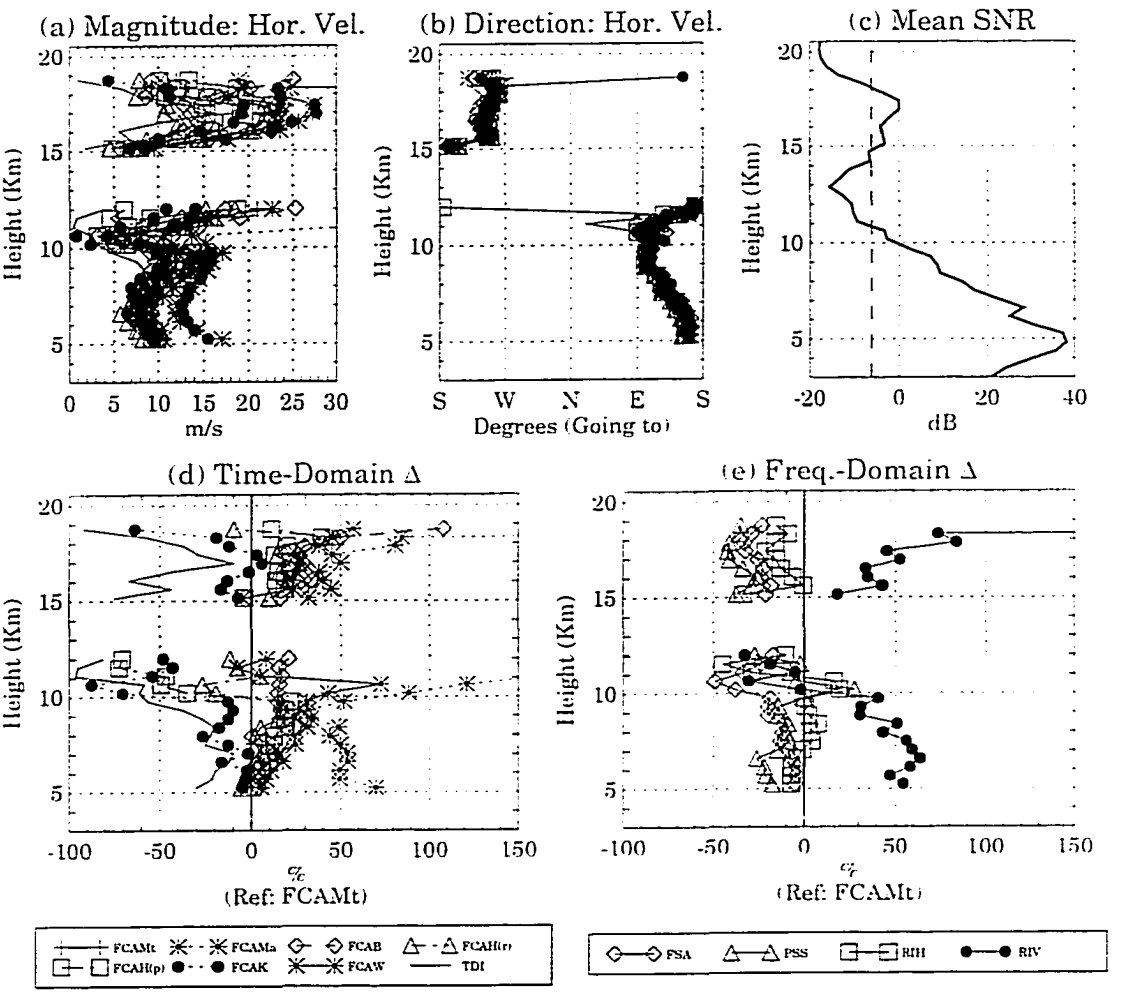


Figure 4.6. Mean horizontal velocity profiles from all techniques for the active period (25-Mar-97 to 29-Mar-97). (a) Magnitude, (b) Direction and (c) Mean SNR profile (the vertical dashed line represents the SNR threshold used to reject data). Relative magnitude difference in percent (Δ) for (d) time-domain techniques and (e) frequency-domain techniques at the bottom. See Section 4.2 for description of the different techniques.

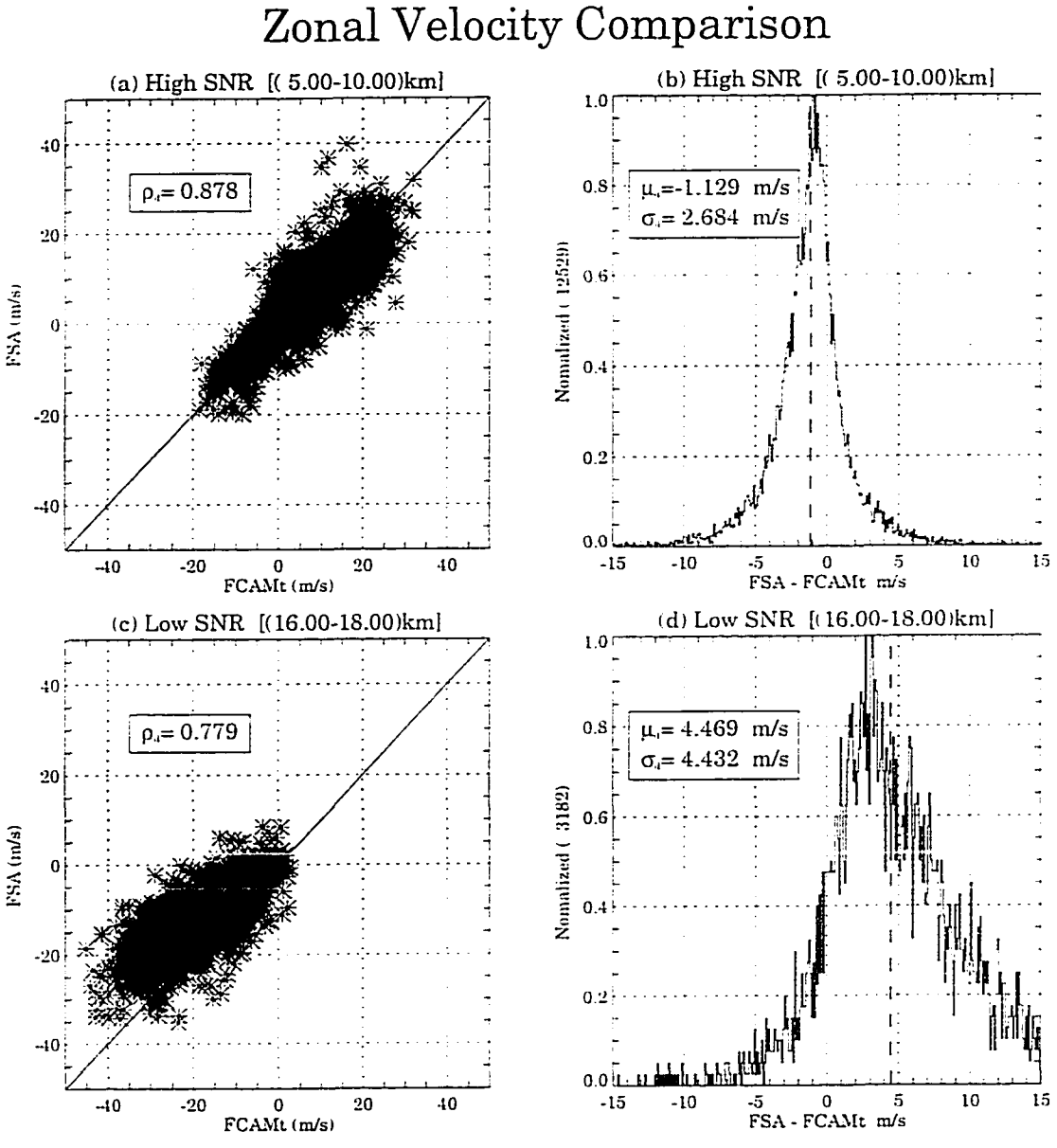


Figure 4.7. Examples of correlation and histogram comparisons of 3-minute zonal velocities obtained by the FCAMt and the FSA SA techniques during the active period. (a) and (b) High SNR (5-10 km), and (c) and (d) Low SNR (16-18 km) comparisons. ρ_d represents the correlation coefficient. μ_d the mean and σ_d the standard deviation of the velocity differences (FCAMt - FSA).

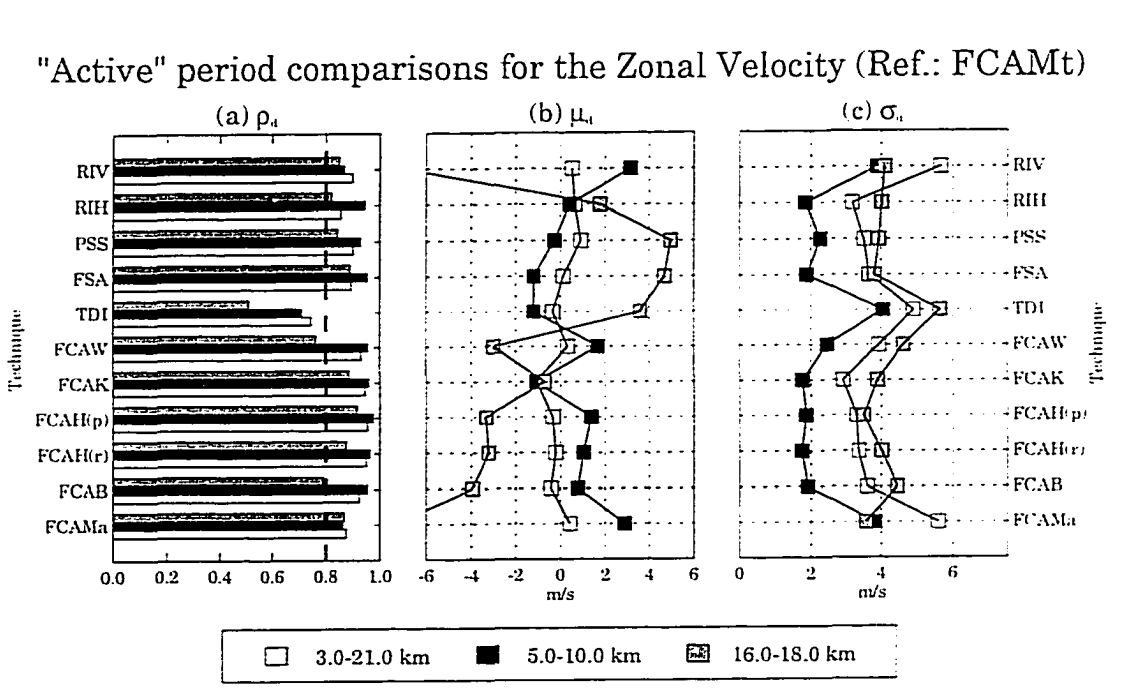


Figure 4.8. Statistical comparisons of hourly-average zonal velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray): during the active period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences.

the comparisons, is consistent for the quiet and active periods. The highest α is seen to occur for FCAW technique. In general the near-tropopause measurements show the highest values of α, as expected with low SNR, with the highest values occurring during the quiet period. This difference could be associated to radar sensitivity differences during the two periods. Most of the "failures" arise in "pre-cleaning". For example, for the 5-10 km region the "pre-cleaning" represents a 16% rejection rate, and the "post-cleaning" the other 6% for the FCAMt failures. Note that TDI results are not presented here due to the different effective sampling time used with this technique.

4.5 Discussion

On the basis of the above comparisons, we found an excellent agreement between the mean wind direction of wind obtained by all of the techniques. The

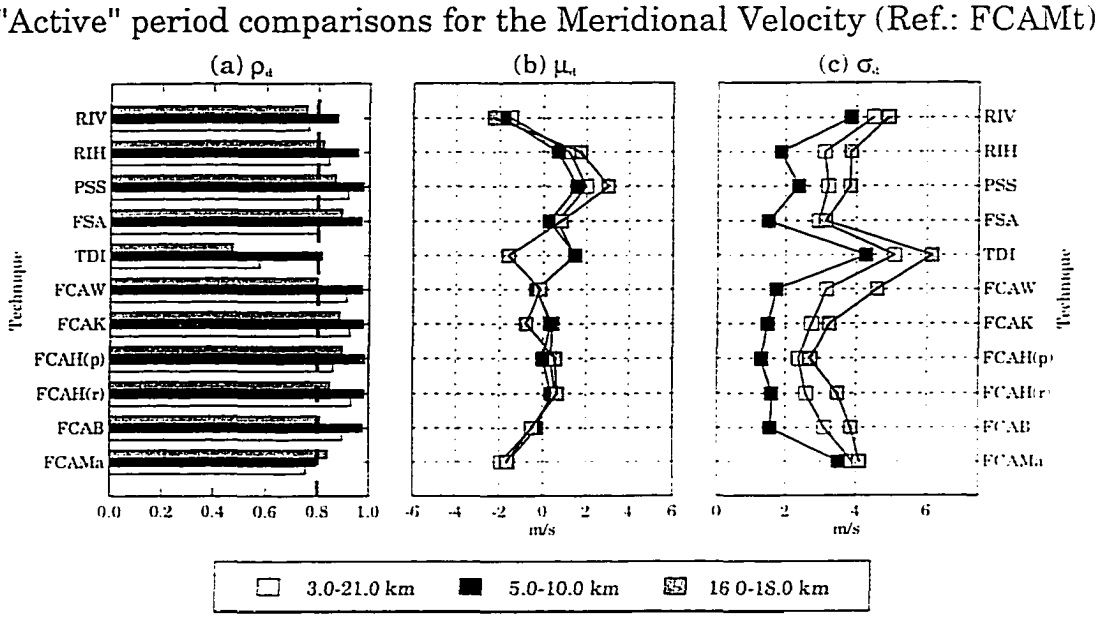


Figure 4.9. Statistical comparisons of hourly-average meridional velocity for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray): during the active period (20-Jan-97 to 27-Jan-97). (a) Correlation coefficient (ρ_d). (b) Mean (μ_d) and (c) standard deviation (σ_d) of the velocity differences.

major discrepancies occur primarily in the wind magnitude. In this section we discuss the degree of agreement of the different techniques and outline possible causes of these discrepancies.

In Section 4.2, we have seen the close relationship between the FCAB and the FCAMt (our reference). Because of this we would expect an excellent agreement in the two techniques. We see this agreement in the high correlation (≈ 0.98), the close-to-zero μ_d and the small σ_d at high SNR heights. The discrepancies, especially at low SNR heights, are primarily due to the greater uncertainties on the estimation of correlation peak positions and crossings values used by the FCAB technique, while the FCAMt have smaller uncertainties with the Gaussian fitting.

The FCAK velocity comparison shows better agreement (higher correlation and less underestimation) during the active period than during the quiet period. This result implies a possible correlation between the accuracy of this technique and the

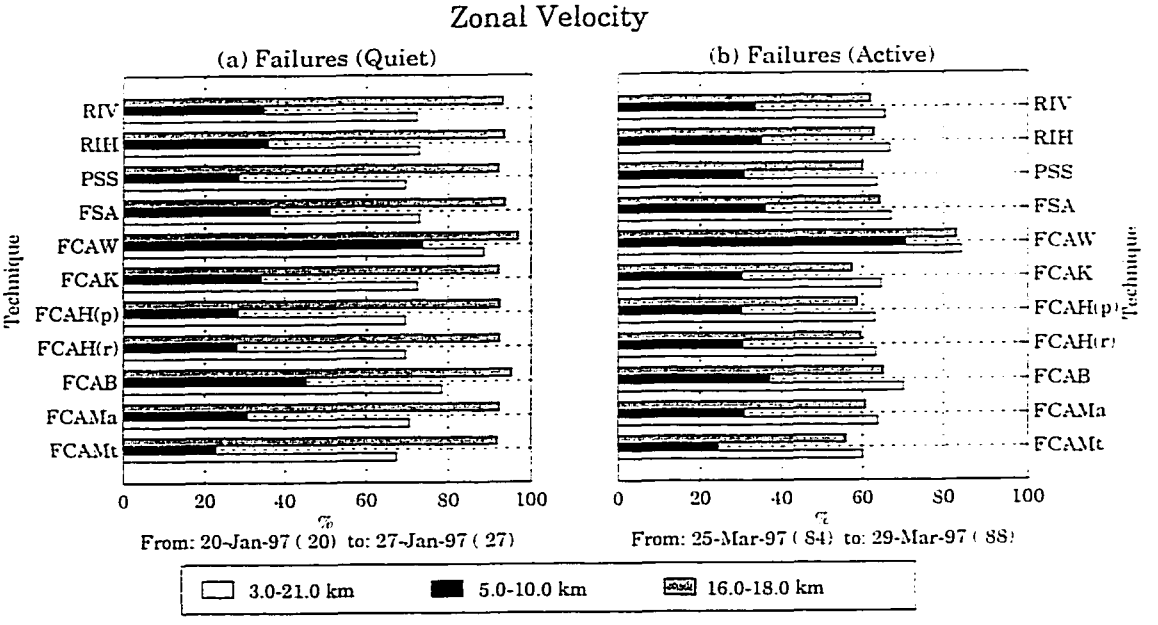


Figure 4.10. Percentage of 3-minute zonal velocity failures (α) for three sets: 3-21 km (light gray), 5-10 km (black) and 16-18 km (dark gray). (a) The quiet period. (b) The active period. This failures represents the estimates rejected in both cleaning stages: "pre-cleaning" and "post-cleaning".

inverse of the width of the ACF (i.e. the correlation time). During the active period the correlation time is shorter and the correlation samples are less correlated with each other, resulting in a better estimation of the wind magnitude.

The three time-domain techniques (FCAW, FCAH(r) and FCAH(p)) that rely on the assumption of horizontally isotropic, i.e. statistically homogeneous, scattering, show excellent agreement with the reference we have chosen (FCAMt)in the troposphere. This is the first time that these techniques have been tested at VHF, although they have been compared at UHF (Cohn et al. [1997]; Holloway et al. [1997a]). From the excellent agreement shown here these techniques are self consistent, at least for high SNR tropospheric heights.

Another way of testing the validity of the FCAH(r) and FCAH(p) measurements, is by comparing the theoretical antenna parameter, a_h, to measured β_h values obtained from Equation 4.6. For isotropic scattering, $\beta_h \approx a_h$ (see Holloway et al.

[1997a]). In Table 4.3, we show mean and standard deviation of β_h values obtained from measured values, for the FCAH(r) and FCAH(p) techniques for both quiet and active periods. These values show excellent agreement with the theoretical value of 0.040. As suggested by Cohn et al. [1997], comparisons of measured β_h values with theoretical values can be used as a "pre-cleaning" criterion.

Table 4.3. β_h results from histograms of the calculated values using Equation 4.6. The theoretical value is 0.040

Technique	Period	
	quiet	active
FCAH(r)	0.039 ± 0.008	0.038 ± 0.011
FCAH(p)	0.041 ± 0.010	0.039 ± 0.012

One advantage of using "isotropic" techniques lies with the small number of parameters required to estimate the horizontal wind. If $\beta_h = a_h$ is assumed, FCAW and FCAH(r) techniques use only two parameters to determine the wind velocity, while the FCAB technique requires four, provided that we assume a horizontally isotropic scattering (i.e., $A = B$ and $H = 0$ in Equation 4.2). Normally the FCAB technique requires six key time lags (see, e.g. Hocking et al. [1989]). Moreover, comparison between the results of FCAH(p), which relies on a Gaussian fit of the CCFs, and those from the FCAH(r), which uses "raw" CCFs, shows excellent agreement at high SNR heights. At near-tropopause heights, FCAH(p) results showed slightly better agreement with the reference than the FCAH(r) results. Therefore, reasonable and fast results could be obtained without a Gaussian fit, i.e. using FCAH(r).

Discrepancies in the FCAW are associated with the high percentage of failures (α), suggesting a poor estimation of the crossings of the CCFs. The crossing estimates depend on the size of the "window" used (e.g. 20% of total lags) and on the correlation time of the signals. An "adaptive" approach, changing the size of the window, may improve the outputs of this technique when a Gaussian form is not assumed (Recall from Section 4.2 that, if a Gaussian form is assumed the FCAW

technique is the same as the FCAH(p)).

The largest discrepancies, for all heights and all conditions, are given by the TDI technique, with lower correlations ($\rho_d \leq 0.8$), large standard deviations of the differences, and considerable underestimations. Vandepeer and Reid [1995] show good agreement with half-hour average FCAB estimates at MF and at mesospheric heights. However, in their configuration they used a wider transmitting beam (10° HPBW), sometimes tilted 10° off-vertical. This configuration includes atmospheric sampling at larger off-vertical angles with correspondingly larger AOAs and radial velocities. The contributions of random errors in the AOAs are also smaller. In our study we used a vertical beam with an $\approx 4°$ HPBW. Most of our AOA measurements were confined to less than 4°, so that random errors in the AOA measurements may be the cause of the differences we present. Other possible causes for these discrepancies, are the effects of antenna spacings and anisotropic scattering on AOA measurements (e.g. May [1993]).

Frequency domain techniques dealing with turbulence, i.e. FSA and RIH show less correlation and a slight underestimation compared with their time-domain counterparts (FCAB and FCAH respectively). Possible factors that could cause these discrepancies include:

- Spectral width estimation. From Equation 4.10 we see that an underestimation of the spectral width results in an underestimated RIH horizontal velocity. This dependence on the spectral width can also be seen by the FSA estimates. An improved spectral width estimate can be obtained by smoothing the CSFs as suggested by Sheppard et al. [1993] or by iteratively eliminating the dominant spikes of the spectra as suggested by Holdsworth [1997]. As mentioned in Section 4.2 we smoothed the spectra in our analysis. Without the smoothing, the velocity underestimation was more severe (not shown) relative to smoothing.
- Linear fitting of the phase of the CSFs. An underestimate of the phase slope of

the CSFs is usually caused by the random phases in the noisy portions of the spectra when a poor fitting is performed. This underestimation of the phase slope, as pointed out by Sheppard et al. [1993] and Holdsworth [1997], produces an overestimation of the "apparent" velocity (RIV). On the other hand, examination of Equation 4.10 suggests that an underestimation of the phase slope produces an underestimation of the RIH velocity. However, there is a good agreement between the time-domain "apparent" velocity (FCAMa) and the frequency-domain "apparent" velocity (RIV) (Figure 4.2 and Figure 4.6). Therefore the poor linear fitting does not appear to be the main source of the FSA and RIH underestimation since RIV measurements rely on the linear fitting of the phase.

• Number of incoherent averages. Incoherent averaging is one way of smoothing the CSFs. The dependence of frequency domain techniques on the number of incoherent averages has being mentioned by Franke et al. [1990] and Holdsworth [1997]. Holdsworth [1997] recommended using a sufficient number of incoherent averages to produce comparable FSA (referred as BVFSA in his paper) and FCAB results. His results show that the errors in the FSA measurements are larger than the FCAB measurements, a result similar to our results discussed in Section 4.4.

• Time and Frequency fitting regions. Sheppard et al. [1993] pointed out that some of the differences between the FCAMt and the FSA techniques may be due to the slightly different temporal scales emphasized by the two domains. FCA type of techniques, concentrate on the lower lags of the CCFs whereas frequency-domain technique (FSA type) place more emphasis on the lower frequency components of the CSFs.

The underestimation in the PSS velocities are primarily due to the off-vertical pointing direction combined with an aspect sensitivity function which biases the synthesized pointing directions. Palmer et al. [1993] also related the underestimation of PSS velocities to aspect sensitivity. However, the high ρ_d (≥ 0.95) and

small σ_d are encouraging results. The advantages of the PSS techniques are related to the advantages of the DBS method and are discussed by Sürücü et al. [1992].

The parameters needed for the majority of time-domain techniques converge to a common set of parameters if we assume a Gaussian form to the CCF magnitudes. Therefore, once the main Gaussian parameters (amplitude, delay and width) have been saved, the following techniques could be implemented: FCAB, FCAM, FCAK, FCAW, FCAH(p). It would also be possible to implement a new approach presented by Holloway et al. [1997b] that includes horizontally anisotropic scattering. Thus, in a sense, the time domain techniques form a set with varying estimators analogous to using different moment estimators with the DBS method.

A similar approach could be applied to the frequency domain techniques, provided that a Gaussian fit is applied to the magnitudes, a linear fit to the phases of the CSFs, and if the amplitude, mean Doppler, spectral width, phase slope and power are saved. Thus not only the FSA, RIV, and RIH techniques could be used, but also some other FSA approaches (e.g. Sheppard and Larsen [1992], Liu and Pan [1993]).

In both domains, when a Gaussian form is assumed, one way to check for the quality of the measurements is to intercompare the widths of the different CCFs/CSFs. In theory, all the widths should be the same as, or close to, the autocorrelation/self-spectrum width (e.g. Meek [1980]; Briggs and Vincent [1992]).

Since CCFs and CSFs are Fourier transform pairs, all the aforementioned techniques (FCAB, FCAM, FCAK, FCAH(p), FSA, RIV, RIH) could be applied provided that a Gaussian form is assumed and the fitting is done just in one domain. The question is: which domain?. One could select the time domain, where CCFs are usually smooth and monotonically decreasing, but the correlation samples are correlated with "neighbor" samples. Alternatively, one could select the frequency domain where spectral estimates are uncorrelated from each other but CSFs are

"noisier".

4.6 Conclusions

In this paper, we have compared the full correlation analysis approach by Meek [1980] (FCAMt) with other time and frequency domain SA techniques. Comparisons have been made with mean values of magnitude, direction, and magnitude difference of the horizontal winds. We have also statistically intercompared 3-minute and hourly values of zonal and meridional velocities. These intercomparisons have been made using correlation coefficients, mean differences, and standard deviation of the differences for all the techniques relative to the FCAMt reference technique. In addition, we have presented the percent of failures for each of the techniques.

Time-domain horizontal velocity measurements techniques that rely on the assumption of statistically isotropic patterns (FCAH(p), FCAH(r), FCAW) show excellent agreement with measurements using more complicated full correlation analysis techniques (FCAM, FCAB). These "isotropic" techniques could represent alternatives for situations where one needs fast measurements of the horizontal wind, particularly at tropospheric heights. Given the excellent agreement between the FCAH(r) and the FCAH(p) results, a Gaussian fit is not recommended when fast measurements are needed.

Frequency domain techniques that correct for turbulence (i.e. techniques that incorporate spectral width) exhibit larger σ_d than their time-domain counterparts. These techniques (FSA and RIH) also consistently underestimate the wind values and require a good spectral width estimation.

The assumption of a Gaussian form of the CCFs is demonstrated to be reasonable at high SNR tropospheric heights, given the excellent agreement of the reference (FCAMt) with techniques not assuming this form (i.e. FCAB and FCAW).

Other possible sources of errors, particularly for the TDI (underestimation

and larger σ_d) and PSS (underestimation), involve the effects of the antenna spacing and anisotropic scattering on AOA measurements and effective pointing positions (e.g. May [1993]). Such effects have not been considered here. They will be subjects for future studies.

The "apparent" velocities (i.e RIV, FCAMa), as expected, show a consistent overestimation ($\approx 50\%$) at all heights. At the very least, their usage is questionable. We do not recommend incorporating the "apparent" velocity measurements as one of the potential techniques for horizontal velocity determination.

In general, for all the techniques examined, there is excellent agreement in the wind direction. On the other hand, considerable discrepancies occur in the wind amplitude estimates. This problem needs further study. Depending on a number of factors, the differences in the horizontal velocity estimates can be significant. For example, errors in estimating the horizontal wind magnitude can produce errors in the correction of vertical wind measurements (e.g. Röttger and Ierkic [1985]; Larsen and Röttger [1991]; Palmer et al. [1991]; Van Baelen et al. [1991]), particularly when the horizontal wind is used in conjunction with angle-of-arrival measurements to get a "true" vertical velocity. Such biases in vertical velocities can produce a major impact in meteorological studies.

Even though the results presented here are based on an arbitrary reference velocity (we have chosen the FCAMt), the majority of the results can be implied if any other reasonable technique were to be chosen. For example, overestimation by "apparent" velocities (FCAMa, RIV), underestimation by FSA, PSS and TDI and the good agreement in the direction of the wind should all be expected.

In the results presented here, the major discrepancies occur near tropopause heights. One can attribute these discrepancies to the percent of failures presented in Figure 4.10 and to the low SNR. Alternatively, a different scattering mechanism may also contribute. In future studies, comparisons at near-tropopause heights should be

done with higher radar sensitivity in order to understand better the differences in this region.

Finally, based on the above discussions, it is impossible to select the best technique. This is due to the fact we do not know the true wind field. Our purpose here has been only to intercompare the major SA techniques that are currently used for horizontal wind measurements.

CHAPTER 5

INTERPRETATION OF ANGLE-OF-ARRIVAL (AOA) MEASUREMENTS IN THE LOWER ATMOSPHERE USING SPACED ANTENNA RADAR SYSTEMS

5.1 Introduction

Atmospheric measurements by very high frequency (VHF) radars tend to show aspect sensitivity effects, i.e., the received signal strength decreases as the beam is tilted off vertical (e.g. Green and Gage [1980]; Tsuda et al. [1986, 1997b]). Usually the aspect sensitivity function is centered at zenith. Some observations, however, show that on occasions the highest reflectivity arrives from small off-vertical locations ($< 2°$) (Green and Gage [1980]; Röttger and Ierkic [1985]; Röttger et al. [1990a]; Tsuda et al. [1997a]). These off-vertical centers have been related to tilted scattering/reflecting structures (e.g. Röttger et al. [1990a]; Larsen and Röttger [1991]; Palmer et al. [1991]). Tsuda et al. [1997a] recently suggested that these off-vertical centers are caused by a corrugated layer that is modified following the vertical displacement caused by gravity waves.

Measurements of these off-vertical centers require systems with sufficient flexibility in beam-pointing and beam-steering capabilities, like the MU radar in Japan (see, Fukao et al. [1990]); or, alternatively, a spaced antenna (SA) system (Larsen and Röttger [1989]; Hocking et al. [1989]).

Most VHF radars currently operate in a Doppler beam swinging (DBS) mode, where the antenna beam is sequentially switched through a series of (at most five) beam positions. Because of the lack of many, closely-spaced, beam positions, measurements of off-vertical centers are not performed with these radars. SA systems, on the other hand, are able of obtaining angle-of-arrival (AOA) returns by

measuring the phase path differences between signal returns using three or more spatially separated, vertically pointing, receiving antennas (Röttger and Ierkic [1985]). These AOA measurements can be associated with off-vertical signal returns.

AOA measurements at VHF, have been used to correct vertical wind measurements by correcting for "leakage" of the horizontal wind into the vertical measurements when off-vertical returns are present (Larsen and Röttger [1991]; Palmer et al. [1991]; Van Baelen et al. [1991]). Recently, Vandepeer and Reid [1995] and Thorsen et al. [1997] have used AOAs to measure the three-dimensional wind field using medium frequency (MF) radars.

The importance of the relative positions of the transmitting and receiving antennas in AOA measurements and vertical velocity corrections, has been pointed out by May [1993]. May also suggested that these measurement should be done with SA systems using "collocated" transmitting and receiving antennas in order to unambiguously interpret the AOA in terms of tilted layers. In addition, according to May, the "leakage" of the horizontal velocity in the vertical beam can be accurately removed only if the antennas are "collocated" (i.e., receiving antennas are symmetrically placed about the transmitter).

In this paper we present the theoretical aspects of various contributions to AOA measurements using SA systems. This development will help to clarify the interpretation of AOA measurements, and their usage in vertical velocity corrections. We have expanded on the scattering model presented by Doviak et al. [1996] (hereafter Doviak et al.), to investigate AOA returns. The modifications to Doviak et al. consist of adding a tilted layer and a slightly off-verticaltransmitting (or receiving) beam, and by explicitly showing the phase terms of Doviak et al.'s original model. We then examine two different methods (time- and frequency-domain) for determining AOAs. We also analyze an expression for the statistical errors in these measurements. AOA measurements using non-collocated SA systems are discussed

in Section 5.3 in terms of vertical velocity corrections and off-vertical atmospheric returns. In Section 5.4, we compare AOA results obtained from both time- and frequency-domain methods. In Section 5.4 we also compare some of the expressions developed in Section 5.2 against experimental data. In addition, a number of interesting time series of orthogonal AOA results are shown. A summary of our results is provided in Section 5.5.

5.2 Angle-of-arrival theory for SA systems

In this section, assuming horizontally isotropic scattering, we first use the scattering theory presented by Doviak et al. to show different contributions to AOA measurements. We then present two different methods of measuring AOAs. Finally, we relate a previously published statistical error for AOA measurements (Thorsen [1996]) with Doviak et al.'s development. Expressions for the horizontal anisotropic case are derived in Appendix E and the results are related to the horizontal isotropic case in this section.

5.2.1 AOA contributions

Here we make use of the formulation derived by Doviak et al., by including some additional off-vertical contributions and by explicitly presenting the important phase terms of the complex cross-correlation functions (CCFs). As pointed out by Röttger and Ierkic [1985], AOAs are obtained from the phase differences between signal returns at receiving antennas. These phase differences can be obtained from the phase terms of the complex CCFs. The end result will be a modified expression of the phase term of the normalized CCF given in Doviak et al.'s Equation 58,

$$\phi_{ij}(\tau) = -2k_0 v_{0z}\tau \tag{5.1}$$

where ϕ_{ij} is the phase term of the CCF between receiving antennas i and j, k_0 is the radar wave number, v_{0z} is the vertical velocity and τ is a time delay. It is important

to point out that Equation 58 in Doviak et al. is valid for collocated SA systems and is a simplify version of their Equation 56.

We consider the following AOA contributions:

• Tilted layer contribution. We have represented a tilted layer effect by incorporating a tilted spatial spectrum of the refractive index field. This tilted spectrum is represented by the following Gaussian function

$$\Phi_n(K) = \Phi_n(0) \exp\left[-\frac{\tilde{\rho}_{ch}^2 |K_h - K_z\delta_0|^2 + \rho_{cz}^2 |K_h|^2 + \tilde{\rho}_{cz}^2 K_z^2}{2}\right] \tag{5.2}$$

where

$$\tilde{\rho}_{ch}^2 = \rho_{ch}^2 - \rho_{cz}^2 \tag{5.3}$$

$$\tilde{\rho}_{cz}^2 = \rho_{cz}^2 - \rho_{ch}^2|\delta_0|^2 \tag{5.4}$$

$$\delta_0 = (\delta_{0x}, \delta_{0y}) \tag{5.5}$$

ρ_{ch} and ρ_{cz} are the irregularity correlation lengths in the parallel and orthogonal directions of the tilted layer, and δ_{0x} and δ_{0y} are the small off-vertical angles of the normal to the tilted plane in the x and y axes, respectively. Note that if $\delta_0 = (0,0)$, Equation 5.2 is equal to Doviak et al.'s Equation 28.

Using Equation 5.2 in Doviak et al.'s Equation 21 and assuming that $\Phi_n(K)$ varies slowly compared to the spectral sampling function $[F_s(K)]$ for K along K_z about the Bragg wavenumber $(2k_0)$, Equation 5.1 becomes

$$\phi_{ij}(\tau) \approx -2k_0 v_{0z}\tau - 4k_0[v_{0h}\cdot\Omega]\xi_h^{-2}\tau + 2k_0[\Delta\rho_{ij}\cdot\Omega]\xi_h^{-2} \tag{5.6}$$

where $[\cdot]$ denotes an inner product operation, v_{0h} is the horizontal vector velocity (v_{0x}, v_{0y}), and

$$\Omega = 2\tilde{\rho}_{ch}^2\delta_0 \tag{5.7}$$

$$\xi_h^2 = 2a_h^{-2} + (2\rho_{ch})^2 \tag{5.8}$$

$$a_h \approx k_0\alpha\sigma_T/z_0 \tag{5.9}$$

$$\alpha^2 = \frac{2\sigma_R^2}{\sigma_T^2 + \sigma_R^2} \tag{5.10}$$

$$\Delta\rho_{ij} = \rho_j - \rho_i \tag{5.11}$$

The parameter α is a function of the difference in the sizes of the transmitting and receiving antennas, ξ_h is related to the horizontal correlation length of the diffraction pattern, a_h is an antenna parameter, z_0 represents the height range under study, $\sigma_{T/R}$ is the transmitting/receiving antenna beam width, and ρ_i/ρ_j is the vector position of receiving antenna i/j. The term $2\Omega\xi_h^{-2}$ is similar to the anisotropic part of Equation 6 presented by Muschinski [1996]. Moreover, this term is proportional to the antenna beam width when the antenna beam width is less than the width of $\Phi_n(\boldsymbol{K})$. The width of $\Phi_n(\boldsymbol{K})$ is $\propto 1/\rho_{ch}$ and is also proportional to the width of the aspect sensitivity function.

This off-vertical contribution could be caused by scattering/reflecting structures (Röttger et al. [1990a]; Palmer et al. [1997b]) that could be associated to Kelvin-Helmholtz instabilities (Muschinski [1996]). Larsen and Palmer [1997] showed that this type of off-vertical AOA can also be produced by gradients in the flow due to divergence.

Examining Equation 5.6, we see that if the scatter is isotropic (i.e. $\rho_{ch} \ll a_h^{-1}$ so $\xi_h^2 \approx 2a_h^{-2}$), then there is no tilted layer contribution ($\Omega = 0$). For non-isotropic scatter, on the other hand, measurements of ξ_h^2 are needed for accurate measurements of δ_0.

• Off-vertical beams. We now add the contribution arising from an off-vertical transmitting beam by assuming that its gain function is represented by,

$$g_T^{1/2}(\boldsymbol{\rho}) = g_T^{1/2}(0) \exp\left[-\frac{|\boldsymbol{\rho} - z_0\boldsymbol{\theta}_0|^2}{4\sigma_T^2}\right] \tag{5.12}$$

where $\boldsymbol{\rho}$ represents the horizontal vector (x, y), σ_T is the transmitting antenna beam width, and $\boldsymbol{\theta}_0 = (\theta_{0x}, \theta_{0y})$ are the small off-vertical pointing angles in the x and y axes, respectively. It is important to point out that the transmitter is located at

(0,0) coordinates in the x-y plane and that this modification does not change the Fresnel zone centers. In addition, we have assumed a non-tilted spatial spectrum of the refractive index field.

Using Equation 5.12 in Doviak et al.'s Equation 10, Equation 5.1 is now modified and represented by,

$$\phi_{ij}(\tau) \approx -2k_0 v_{0z}\tau - 4k_0[v_{0h}\cdot\Psi]\xi_h^{-2}\tau + 2k_0[\Delta\rho_{ij}\cdot\Psi]\xi_h^{-2} \tag{5.13}$$

where

$$\Psi \approx \frac{\alpha^2\theta_0}{2a_h^2} \tag{5.14}$$

In Equation 5.13, note the dependence on ρ_{ch} via ξ_h^2. For a horizontally isotropic atmosphere, i.e., $\xi_h^2 \approx 2a_h^{-2}$, this contribution will depend just on the pointing direction of the antennas, the antenna beam widths, and the height range.

In the case the receiving beams are pointing off-vertical, instead of the transmitting beam, Ψ is given by

$$\Psi \approx \frac{\alpha^2\sigma_T^2\theta_0}{2\sigma_R^2 a_h^2} \tag{5.15}$$

where θ_0 is the off-vertical pointing angle of the receiving antennas, and the following function was assumed for the receiving pattern

$$g_R^{1/2}(\rho) = g_R^{1/2}(0)\exp\left[-\frac{|\rho-(\rho_i - z_0\theta_0)|^2}{4\sigma_R^2}\right] \tag{5.16}$$

Off-vertical pointing beams could be due to small imperfections on vertically-directed antennas (Huaman and Balsley [1996]) or to intentional tilting of the antenna beams. For example, Vandepeer and Reid [1995] used off-vertical beam directions (10°) in their SA configuration, to obtain the wind velocity vector from AOA measurements.

• Geometrical contribution dependent on the horizontal correlation scale. For this third contribution, we have again assumed a non-tilted spatial spectrum of the refractive index field. By grouping geometrical phase terms that depends on ρ_{ch}, Equation 5.1 is modified into,

$$\phi_{ij}(\tau) \approx -2k_0 v_{0z}\tau - 4k_0[v_{0h}\cdot\Gamma_{ij}]\xi_h^{-2}\tau + 2k_0[\Delta\rho_{ij}\cdot\Gamma_{ij}]\xi_h^{-2} \tag{5.17}$$

where

$$\Gamma_{ij} \approx -\frac{(\rho_i + \rho_j)(\alpha^2 - 1)}{4z_0 a_h^2} \tag{5.18}$$

Examining Equation 5.18, Γ is a function of the geometry of the system, i.e., of the antenna positions and of the antennas sizes (α), and decreases with increasing height range. However, the whole contribution varies as ρ_{ch} changes (via ξ_h^2 in Equation 5.17). For example, independent of the value of ρ_{ch}, this geometric contribution becomes zero: (a) when the receiving antennas i and j are located symmetrically respect to the transmitting antenna ($\rho_i = -\rho_j$), or (b) when all the antennas have the same size, i.e., the same beam widths ($\alpha = 1$). Note that for an auto correlation function ($i = j$) and if $v_{0z} = 0$, $\partial\phi_{ii}(\tau)/\partial\tau$ reduces to Doviak et al.'s Equation 36.

• Geometrical contribution independent of the horizontal correlation scale. This last contribution comes from the phase terms that do not depend on ρ_{ch}. Moreover, the contribution of these phase terms, is constant for a specific system's geometry and for a specific height. Including these geometrical phase terms, Equation 5.1 becomes,

$$\phi_{ij}(\tau) \approx -2k_0 v_{0z}\tau + k_0|\Delta\rho_{ij}|\Lambda_{ij} \tag{5.19}$$

where,

$$\begin{aligned} \Lambda_{ij} &\approx -\frac{2(D_j - D_i)}{|\Delta\rho_{ij}|} \\ D_i^2 &= z_0^2 + |\rho_i/2|^2 \\ D_j^2 &= z_0^2 + |\rho_j/2|^2 \end{aligned}$$

This geometrical contribution becomes zero in SA systems where the receiving antennas are equidistant from the transmitter ($|\rho_i| = |\rho_j|$), independent of ρ_{ch} or the pointing direction of the antennas.

It is important to mention that these four AOA contributions do not affect the magnitude of the normalized CCFs. In addition, the two geometric terms are present in Doviak et al.'s Equation 56, although they are not shown explicitly. Moreover, the expressions just presented are valid under the condition $|\rho_i|, |\rho_j| \leq \sigma_T$.

Putting together all the aforementioned contributions, the phase term of the complex CCF, $c_{ij}(\tau)$, is

$$\phi_{ij}(\tau) \approx -2k_0 v_{0z}\tau - 2k_0[v_{0h} \cdot \Theta_{ij}]\tau + k_0[\Delta\rho_{ij} \cdot \Theta_{ij}] + k_0|\Delta\rho_{ij}|\Lambda_{ij} \qquad (5.20)$$

where

$$\Theta_{ij} = 2(\Omega + \Psi + \Gamma_{ij})\xi_h^{-2} \qquad (5.21)$$

so Θ_{ij} is an AOA that includes the tilted layer contribution, the off-vertical beam direction, one geometrical contribution, and the horizontal correlation length of the scatterers (via ξ_h^2).

We have derived this phase expression assuming horizontally isotropic scattering (i.e., $\rho_{ch} = \rho_{cx} = \rho_{cy}$). It is worthwhile to note, that a similar expression can be obtained if we assume horizontally anisotropic scattering (see Equation E.4). In both scattering cases, measurements of the scattering parameters are needed to estimate most AOA contributions. The scattering parameters can be calculated from a FCA analysis (Briggs [1984]; Meek [1980]; Holloway et al. [1997b]).

The need for correction of vertical velocities for horizontal "leakage" can be seen by examining Equation 5.20. Moreover, the three-dimensional velocity vector can be obtained from this equation. This approach has been used by Vandepeer and Reid [1995] and Thorsen et al. [1997] and is called time-domain interferometry (TDI).

In this derivation, we have assumed that the phase difference among receivers lines is zero. Usually these phase differences are not zero (e.g. different cable lengths), and calibration procedures, like the one described in Section 5.4, are needed to removed them. Other calibration procedures have been described by Palmer et al. [1996]; Vandepeer and Reid [1995]; Thorsen et al. [1997]; Röttger et al. [1990a].

5.2.2 Methods of estimating AOAs

Basically there are two methods to estimate AOA returns, a time-domain and a frequency-domain method. A description of these two methods follows:

• Time-Domain method. This method was first proposed by Röttger and Ierkic [1985]. Basically, AOA measurements are obtained by measuring the phase angles of the CCFs at $\tau = 0$ for at least two antenna baselines. From Equation 5.20, the phase of baseline $i - j$ at $\tau = 0$ is,

$$\phi_{ij}(0) \approx k_0[\Delta\rho_{ij} \cdot \Theta_{ij}] + k_0|\Delta\rho_{ij}|\Lambda_{ij} \tag{5.22}$$

then measuring the phase of another baseline at $\tau = 0$, knowing the positions vectors (e.g. ρ_1, ρ_2 and ρ_3), and calculating the constant geometric terms (for instance, Λ_{12}, Λ_{13} and Λ_{23}, for a three receiver SA system), we can calculate the two dimensional AOA Θ.

• Frequency-Domain method. This method was discussed by Briggs and Vincent [1992] for the two dimensional case. Holdsworth [1997] presented an extension for the three dimensional situation. Basically, Fourier transforming Doviak et al.'s Equation 58 and including Equation 5.20, the phase of the resulting cross-spectrum function (CSF) is given by,

$$\phi_{ij}(f) = m_{ij}(f - f_{dij}) + k_0[\Delta\rho_{ij} \cdot \Theta_{ij}] + k_0|\Delta\rho_{ij}|\Lambda_{ij} \tag{5.23}$$

where

$$f_{dij} = -\frac{2}{\lambda}(v_{0z} + [v_{0h} \cdot \Theta_{ij}]) \tag{5.24}$$

is the Doppler shift, λ is the radar's wavelength, and m_{ij} is a slope that depends on v_{0h}, ξ_h^2, $\Delta\rho_{ij}$ and the turbulence intensity (σ_t).

From Equation 5.23, the two-dimensional AOA Θ, can be obtained by evaluating the phase $\phi(f)$ at the frequencies $f = f_d$, and by calculating the geometrical terms Λ, for at least two antenna baselines.

5.2.3 Statistical errors in time-domain AOA measurements

Thorsen [1996] presented an expression for the variance of the estimate of the CCF phase ϕ, assuming the following CCF.

$$C_{ij}[\tau] = S_{ij}\exp\{-8[\pi\sigma_v T_s(\tau-\tau_{pij})/\lambda]^2\}\exp(j\omega_d\tau T_s+\varphi_{ij}) + N_{ij}\delta[\tau] \qquad (5.25)$$

where σ_v is the spectral width, T_s is the sampling time. τ_{pij} is the delay to the peak of C_{ij}, ω_d is a Doppler shift, φ_{ij} a phase difference, S_{ij} is the signal power and N_{ij} is the noise power. The phase difference φ_{ij}, and the delay to the peak τ_{pij}, are equal to zero for $i = j$. The separate receiver noises are assumed uncorrelated ($N_{ij} = 0$ for $i \neq j$).

Instead of using Equation 5.25, we have used Doviak et al.'s Equation 58 along with Equation 5.20. Then, the CCF is given by

$$C_{ij}[\tau] = S_i S_j A_{ij}\exp\left[-\frac{(\tau-\tau_{pij}/T_s)^2}{2(\tau_c/T_s)^2}\right]\exp(j\phi_{ij}[\tau]) + N_{ij}\delta[\tau] \qquad (5.26)$$

where

$$\begin{aligned}
\tau_c^2 &= \frac{1}{2[2\xi_h^{-2}|v_{0h}|^2 + (2k_0\sigma_t)^2]} \\
\tau_{pij}^2 &= 2\xi_h^{-2}(v_{0h}\cdot\Delta\rho_{ij})\tau_c^2 \\
A_{ij} &= \exp\left(-\frac{|\Delta\rho_{ij}|^2}{2\xi_h^2}+\frac{\tau_{pij}^2}{2\tau_c^2}\right)
\end{aligned}$$

Examination of Equation 5.25 and Equation 5.26 shows that they are very similar. However Equation 5.26 relates its amplitude (A_{ij}), delay to the peak and correlation time (τ_c), to the SA parameters (antenna positions, antenna beam widths, etc.) and to the atmospheric parameters (horizontal velocity, turbulence intensity and ρ_{ch}).

Using Equation 5.26 and following the derivation of Thorsen [1996], the variance of an estimated phase is,

$$E[(\delta\hat{\phi}_{ij}[\tau])^2] \approx \frac{1}{2M\mathrm{coh}_{ij}^2[\tau]}\left\{\left(\frac{N_i^2}{S_i^2}+\frac{N_j^2}{S_j^2}+\frac{N_i^2}{S_i^2}\frac{N_j^2}{S_j^2}\right)+(1-\mathrm{coh}_{ij}^2[\tau])\frac{\sqrt{\pi}\tau_c}{T_s}\right\} \qquad (5.27)$$

for $i \neq j$, where

$$\mathrm{coh}_{ij}^2[\tau] = A_{ij}^2 \exp\left[-\frac{(\tau - \tau_{pij}/T_s)^2}{(\tau_c/T_s)^2}\right] \tag{5.28}$$

$\delta\hat{\phi}_{ij} = \hat{\phi}_{ij} - \phi_{ij}$, $\hat{\phi}_{ij}$ are the measured values, ϕ_{ij} represent the true values, and M represent the number of points used in the estimate. For small τ, $E[\hat{\phi}_{ij}] = E[\phi_{ij}]$ (Thorsen [1996]), therefore $E[\delta\hat{\phi}_{ij}] = 0$.

For the time-domain method, the variance of the estimated phase (statistical error), is obtained by evaluating Equation 5.28 at $\tau = 0$. At high signal-to-noise ratio ($\mathrm{SNR}_i = S_i^2/N_i^2$), this variance becomes,

$$\varepsilon_{\phi ij}^2 = E[(\delta\hat{\phi}_{ij}[0])^2] \approx \frac{1}{2M}\frac{(1 - \mathrm{coh}_{ij}^2[0])}{\mathrm{coh}_{ij}^2[0]}\frac{\sqrt{\pi}\tau_c}{T_s} \tag{5.29}$$

where $\mathrm{coh}_{ij}^2[0] = \exp[-|\Delta\rho_{ij}|^2/(2\xi_h^2)]$. Note that this derivation is valid for horizontally isotropic scattering. In the case of horizontal anisotropy, i.e. $\rho_{cx} \neq \rho_{cy}$, Equation 5.29 remains the same. However, the coherence value $\mathrm{coh}_{ij}[0]$, is given by Equation E.9, and the new correlation time by Equation 28 in Holloway et al. [1997b]. In practice, $\mathrm{coh}_{ij}^2[0]$ is calculated after the "spike" at zero-lag, i.e., noise, has been removed from autocorrelation and cross-correlation functions (Briggs [1984]).

5.3 On the use of non-collocated SA systems for AOA measurements

In a collocated SA system, the geometrical center of the transmitting antenna coincides with the geometrical center of the array of receiving antennas. Examples of these systems are equilateral triangles and square configurations, where the transmitting antenna is located in the center of such configurations. In a non-collocated SA system such geometrical centers do not coincide. In this section, we examine some issues on the AOAs obtained by non-collocated SA systems.

Examining Equation 5.20, the only terms that contribute to the measured AOA and depend on the antenna positions are the geometrical terms Γ and Λ. A similar contribution has been pointed out by Vincent et al. [1987] and May [1993]

(hereafter May). Based on this contribution, May suggested that collocated transmitting and receiving antennas should be used in atmospheric AOA and corrected vertical velocity measurements. In the following paragraphs we analyze these assertions.

To begin, we examine a similar example to the one given by May. A narrow transmitting beam and wide receiving beams are assumed, implying $\alpha^2 = 2$. An isotropic scatter is represented with $\xi_h^2 \approx 2a_h^{-2}$, and a specular scatter with $\xi_h^2 \to \infty$. The positions of the receivers, with respect to the transmitter position, are $\rho_1 = (2x_1, 0)$ and $\rho_2 = (2x_2, 0)$. We assume zero vertical velocity ($v_{0z} = 0$), perfect vertically-directed antennas [$\theta_0 = (0, 0)$], and atmospheric contribution centered at zenith [$\delta_0 = (0, 0)$]. With these assumptions we analyze the following two effects:

• Horizontal "leakage". The "leakage" of the horizontal velocity in the radial velocity can be seen by using Equation 5.20 and evaluating the Doppler shift $d\phi_{11}(\tau)/d\tau$. The results for the isotropic and specular cases are,

$$\frac{d\phi_{11}(\tau)}{d\tau} = \begin{cases} \frac{k_0(2x_1)v_{0x}}{z_0} & \text{Isotropic} \\ 0 & \text{Specular} \end{cases} \tag{5.30}$$

which are the same to those presented by May.

• Receiver's separation. Evaluating Equation 5.20 at $\tau = 0$, the phase term $\phi_{12}(0)$ becomes

$$\phi_{12}(0) = \begin{cases} \frac{2k_0(x_1^2 - x_2^2)}{z_0} & \text{Isotropic} \\ \frac{k_0(x_1^2 - x_2^2)}{z_0} & \text{Specular} \end{cases} \tag{5.31}$$

where we have used a Taylor series expansion for D_1 and D_2. Again, these expressions are the same to those presented by May.

Evaluating the Doppler shift from

$$f_d = \left[\frac{d\phi_{11}(\tau)}{d\tau} + \frac{d\phi_{22}(\tau)}{d\tau}\right] \Big/ 2 \tag{5.32}$$

show that the leakage of the horizontal velocity can be removed using the cross-correlation phase $[\phi_{12}(0) - k_0|\Delta\rho_{ij}|\Lambda_{ij}]$, without knowing ρ_{ch} (recall that Λ_{ij} are constant for a given height and depend just on the receiver spacing).

Based on this example we see that May's horizontal "leakage" and receiver's separation effects are related through Equation 5.17, and vertical velocity corrections are also possible using non-collocated systems.

One disadvantage of using non-collocated systems appears in some calibration procedures. In Section 5.2, we saw that a calibration procedure is needed to eliminate the phase difference among receiver lines. Some of these procedures use long-term atmospheric data, and assume that the mean off-vertical atmospheric return is zero (Röttger et al. [1990a]; Thorsen et al. [1997]). However, in the lower atmosphere, such a procedure will calibrate out the geometrical term (Γ) that is needed for vertical velocity corrections.

Following our analysis, the unambiguous determination of the off-vertical atmospheric returns requires accurate measurements of the scattering parameters, i.e., ξ_h^2 when horizontal isotropic scattering is assumed, or ξ_x^2, ξ_y^2 and ψ (orientation) for horizontal anisotropic scattering. The latter parameters can be obtained from a FCA procedure. This requirement is true not only for non-collocated SA systems but also for those that are collocated. On the other hand, vertical velocity corrections can be done without knowing the scattering parameters under both collocated and non-collocated SA systems.

5.4 Experimental results

We have used the large VHF radar at the Jicamarca Radio observatory in Perú for our AOA studies. The Jicamarca radar operates at ≈ 50 MHz ($\lambda \approx 6m$) and has an $\approx$ 300 m by 300 m antenna array. The antenna is composed of 64 separate modules. Each module consists of a 12 by 12 array of cross-polarized half-wave

dipoles. In Figure 5.1, we show the antenna configuration for these experiments. For transmission, we phased four modules (section D) to point precisely vertically, with a one-way half-power beam width (HPBW) of $\approx 4°$. Antenna sections A, B and C were used for reception. Each receiving antenna was pointed on-axis ($\approx -1.46°$ from vertical along the the y-axis, i.e. toward the SW) and had a one-way HPBW of $\approx 8.2°$. The output from each of the receiving antennas was independently received, sampled, and recorded. Ground clutter effects in each receiver were reduced separately in both the real and imaginary signal channels from the recorded data. To do this, we calculated the mean complex voltages in each of the individual time blocks defined by the incoherent averaging period. We then fitted a second degree polynomial to the time-series of the real and imaginary components of these mean values. Ground clutter effects were removed by subtracting this function from the original voltages values to provide the modified time series for subsequent analysis.

We present here the results of a 5-day run taken in March, 1997 (from March 25, 16:50 LT to March 29, 22:15 LT). The operational parameters for this data set are presented in Table 5.1.

Table 5.1. Operating parameters for the experiments conducted at Jicamarca (from 25-Mar-97 to 29-Mar-97)

Parameter	Value
Pulse repetition period	2 ms
Number of coherent integrations	128
Number of incoherent integrations	6
Number of points	128
Effective sampling time	256 ms
Height resolution	0.45 km
Initial height	3 km
Number of heights	40
Transmitting power	100 kW

A calibration procedure was performed before and after the experiment to determine the possible phase differences arising from differing lengths of the receiving

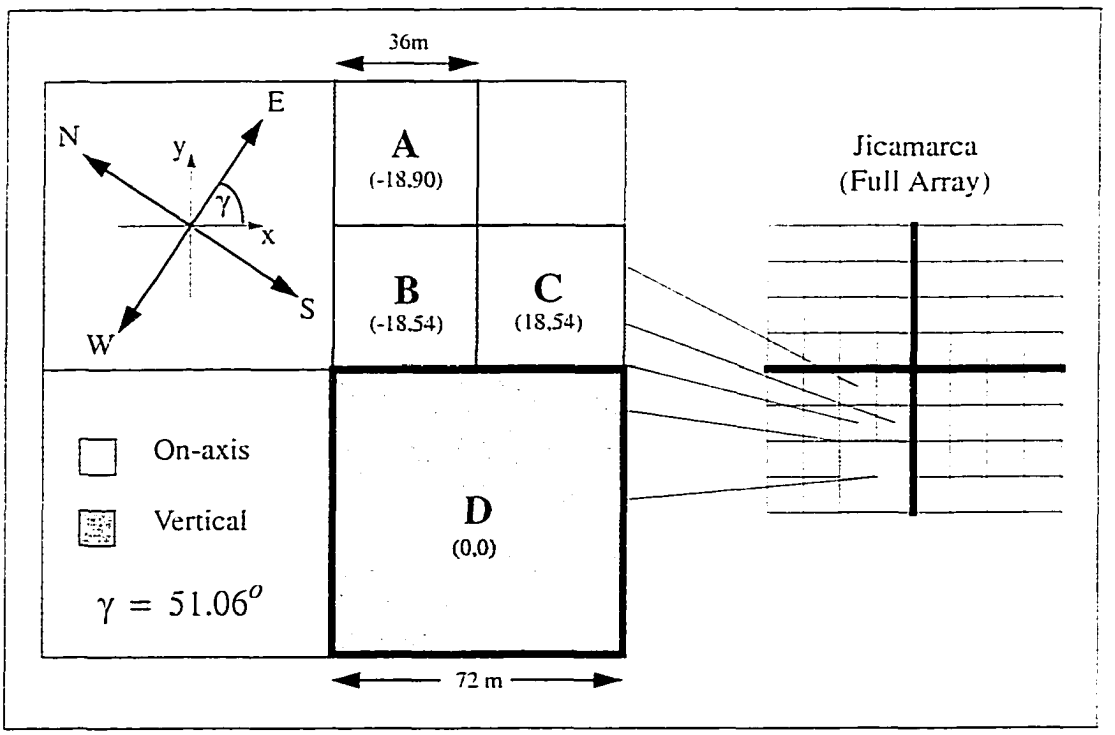


Figure 5.1. Antenna setup for the Jicamarca experiment. The transmitting antenna is the vertically pointing D module: and the receiving antennas are the "on-axis" pointing A, B and C modules. The "on-axis" position differ from the zenith position because antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest. Notice that neither of the antenna axes, (x, y), is aligned with the geographic coordinates (E, N). The antenna centers are given in (x, y) coordinates, respect to the transmitting antenna center.

lines. Briefly, a common signal (either an output from one of the antennas or amplified sky noise) was fed in parallel to each of the receiving lines (cables, front-ends, amplifiers and receivers). This signal was recorded and processed using parameters similar to those listed in Table 5.1. Phase differences determined from this process were incorporated into the subsequent analysis.

AOA measurements were obtained every ≈ 3 minutes using both time- and frequency-domain methods (see Section 5.2.2). In both methods the measured AOA includes the four contributions described in Section 5.2.1. We have ignored estimates with very low signal-to-noise ratio (SNR < -6 dB, after coherent and incoherent integrations) in our analysis. It is important to point out that the AOAs

presented next are relative to the antenna surface and not to the horizontal plane (see Figure 5.1 for details).

We determined the time domain AOAs by fitting a line to the phase of each CCF pair (BA, BC and CA) and evaluating that line at $\tau = 0$. This process was used in order to minimize the residual contributions from noise at $\tau = 0$.

In Figure 5.2, we present histograms of AOAs obtained in this manner in the two orthogonal bases BC and BA (see Figure 5.1) for heights between 7 and 8 km. This figure shows the distributions of "raw" AOA values (solid line) and a Gaussian-fitted function (stars). Mean (μ_Θ) and standard deviation (σ_Θ) values are shown for both the "raw" (subscript r) and Gaussian-fitted (subscript g) data. Note that the results are almost the same with and without Gaussian fitting.

As mentioned above, these AOA values are relative to the antenna baseline, i.e., the true zenith position translates into $\Theta_{BC} = 0°$ and $\Theta_{BA} = 1.46°$.

In Figure 5.3, we present results for the frequency-domain AOAs in a format similar to Figure 5.2 for the same time period. We obtained these measurements by fitting a line to the phase of the CSFs and evaluating them at $f = f_d$, where f_d is the Doppler shift of the respective CSF (see Section 5.2.2). Examination of this figure shows the presence of more "outliers" in the frequency-domain results than using the time-domain method. The "outliers" are graphically manifested by the presence of a constant "floor" extending across the histograms, i.e., values that raise the level of the base of the Gaussian-like functions at all values of AOAs. These "outliers" are also noticed in the standard deviation values, where $\sigma_{\Theta r}$ is considerably larger than $\sigma_{\Theta g}$ for both baselines. Note that the Gaussian fitted values are similar to those obtained from the time-domain measurements.

From these similarities, we see that both methods give essentially the same information. It is important to point out that implementation is easier and calculation is faster in the time-domain. Frequency-domain methods, however, have the

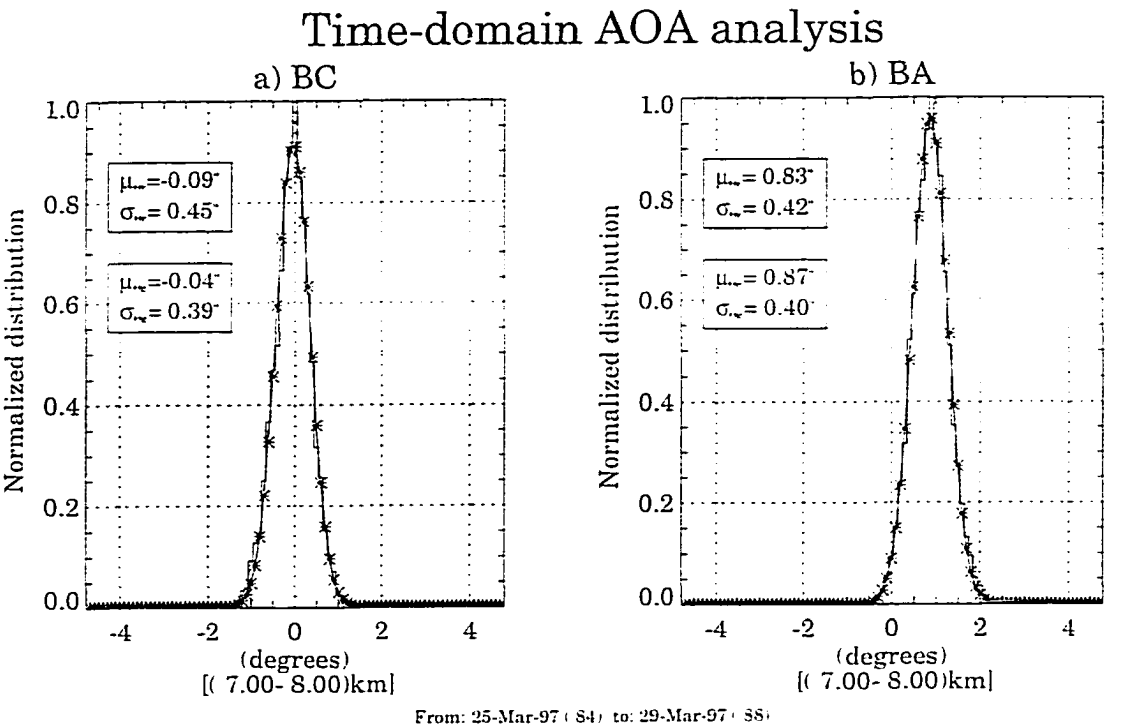


Figure 5.2. Histograms of 3-minute AOAs obtained by a time-domain method (see Section 5.2.2) using the BC (Θ_x), and BA (Θ_y) antenna pairs. Solid lines represent the actual measurements, and the fitted Gaussian functions are represented by stars. Mean (μ_Θ) and standard deviation (σ_Θ) of the AOA measurements are given for the "raw" (subscript r) and Gaussian fitted (subscript g) data. These AOA measurements are relative to the antenna plane (see Figure 5.1) and include heights between 7 and 8 km.

distinct advantage when interference is encountered. In the following analysis, we will use only AOAs obtained with the time-domain method.

We now compared results obtained using the expressions presented in Section 5.2.1 with those obtained from measured values of AOAs. The (x, y) receiver locations relative to the transmitter are $\rho_A = (-18\text{ m}, 90\text{ m})$, $\rho_B = (-18\text{ m}, 54\text{ m})$ and $\rho_C = (18\text{ m}, 54\text{ m})$ (see Figure 5.1). The transmitting beam position is $\theta_0 = (0, 1.46°)$ where 1.46° is the tilt of the antenna plane. The beam widths $\sigma_{T,R}$ (in meters) are related to the $\text{HPBW}_{T,R}$ (in radians) by, $\sigma_{T,R} = z_0 \text{HPBW}_{T,R}/(2\sqrt{2\ln 2})$. Using these values we have calculated the theoretical AOAs assuming an isotropic atmosphere ($\rho_{ch} \ll a_h^{-1}$). These theoretical curves are shown in Figure 5.4 for the

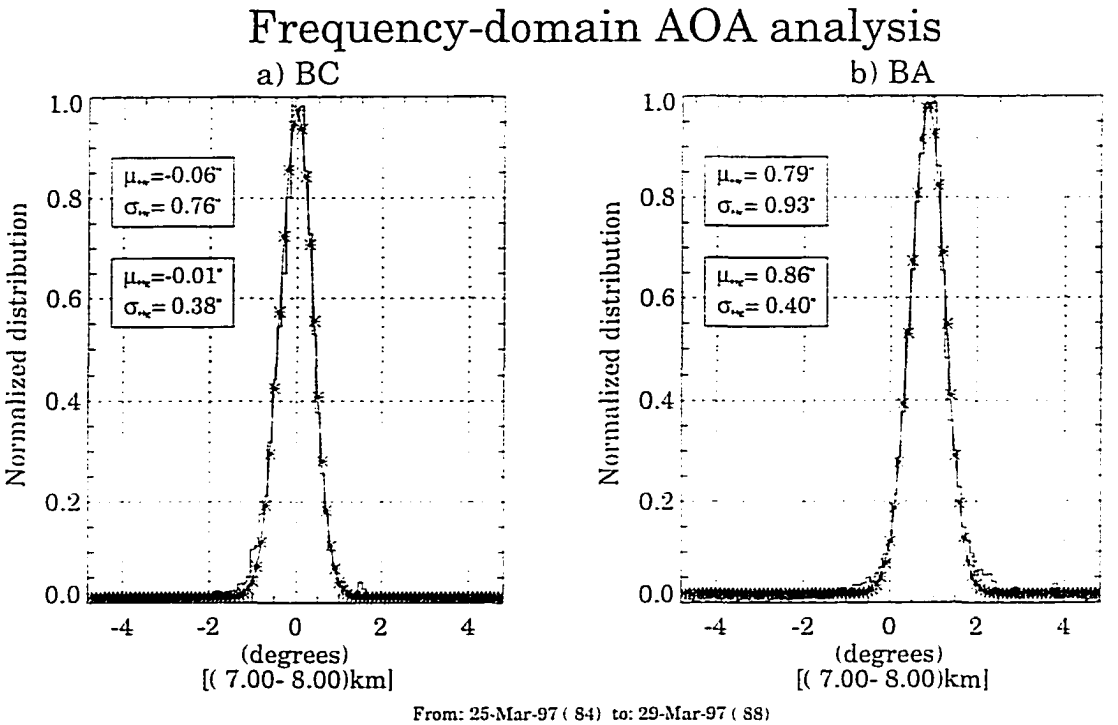


Figure 5.3. Histograms of 3-minute AOAs obtained by a frequency-domain method (see Section 5.2.2) using the BC (Θ_x), and BA (Θ_y) antenna pairs. Solid lines represent the actual measurements, and the fitted Gaussian functions are represented by stars. Mean (μ_Θ) and standard deviation (σ_Θ) of the AOA measurements are given for the "raw" (subscript r) and Gaussian fitted (subscript g) data. These AOA measurements are relative to the antenna plane (see Figure 5.1) and include heights between 7-8 km.

two orthogonal baselines. The transmitting beam contribution ($2\Psi\xi_h^{-2}$) is shown by stars, the variable geometrical contribution ($2\Gamma\xi_h^{-2}$) by pluses, the constant geometrical contribution (Λ) by diamonds, and the combined contribution $[2(\Psi+\Gamma)\xi_h^{-2}+\Lambda]$ by triangles. Notice the constant value of zero, for all contributions, in baseline BC (aligned with x). These values will be the same if the isotropic condition is not satisfy ($\xi_h^2 \neq 2a_h^{-2}$). This result is mainly due to symmetric locations (in x) of receiving antennas B and C with respect to the transmitting antenna (see Figure 5.1).

For baseline BA the combined contribution have an asymptotic behavior with height, towards the antenna pointing contribution ($2\Psi\xi_h^{-2}$). If the atmosphere is specular ($\xi_h^2 \to \infty$), extreme case, the combined contribution will be just the

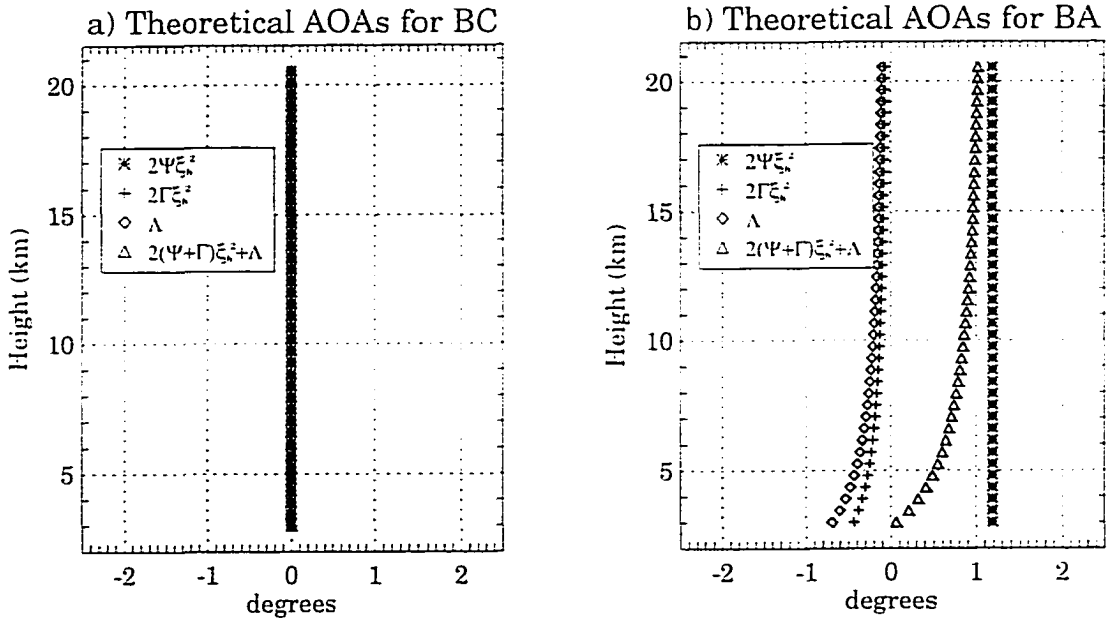


Figure 5.4. Theoretical AOAs under isotropic conditions ($\xi_h^2 \approx 2a_h^{-2}$), including just the SA system contributions. Transmitting beam contribution ($2\Psi\xi_h^{-2}$), two geometric terms ($2\Gamma\xi_h^{-2}$ and Λ), and the combined contribution $[2(\Psi+\Gamma)\xi_h^{-2}+\Lambda]$. (a) BC ($\Theta_x$), and (b) BA ($\Theta_y$).

constant geometrical term (Λ). Moreover, Λ decreases with increasing height.

In Figure 5.5, we present the 5-day AOA measurements in panels (a) and (b), for both orthogonal baselines (*BC* and *BA*), and the mean SNR profile in panel (c). The mean AOAs (triangles) are compared to the theoretical AOAs assuming isotropic conditions (thick line. $[2(\Psi+\Gamma)\xi_h^{-2}+\Lambda]$ shown in Figure 5.4). Sampled standard deviations are denoted with horizontal dashed lines. On the upper axis, we present the median value of the statistical errors of the AOAs ε_Θ (diamonds, see Equation 5.29), with values $< 0.2°$ for heights with good SNR. Notice the excellent agreement between the empirical and theoretical mean values, particularly, in the lower heights ($<$ 15 km). The small differences in the near-tropopause heights ($\approx$ 17 km) could arise from non-isotropic scattering and/or "off-vertical" atmospheric returns $[\delta_0 \neq (0,0)]$ under anisotropic scattering.

We have tried to quantify the accuracy of Equation 5.29 for the statistical

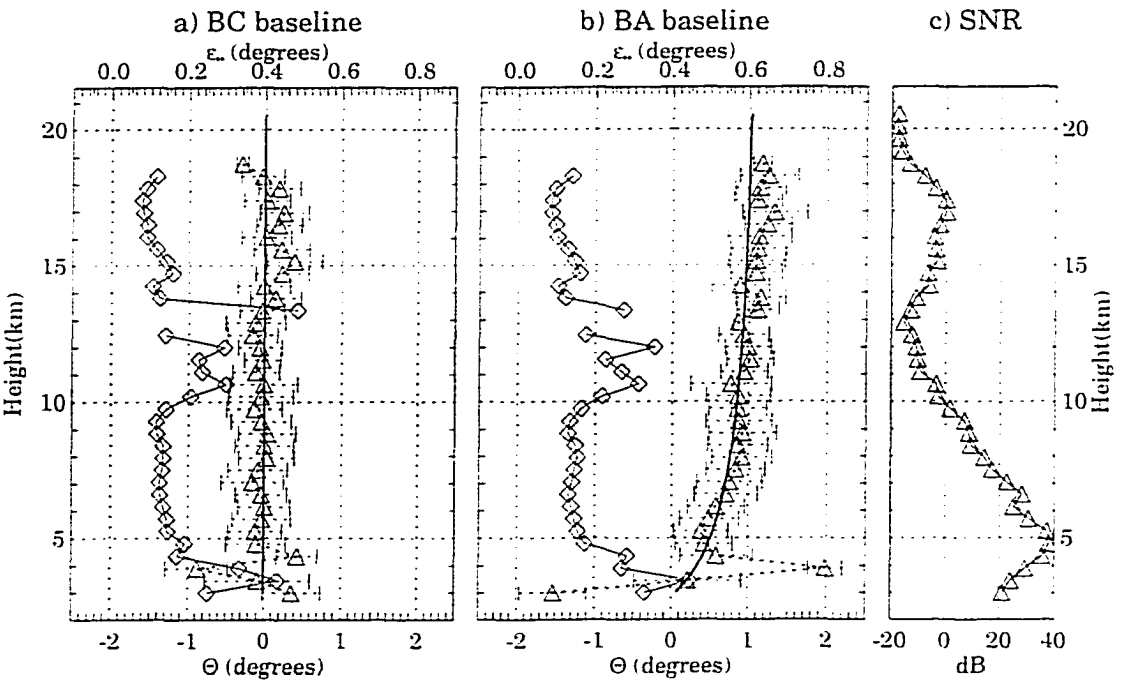


Figure 5.5. AOA profiles for the five-day data set. AOA results for (a) baseline BC (Θ_x) and (b) baseline BA (Θ_y). The mean SNR profile appears in (c). The means and standard deviations of empirical values (triangles and horizontal dotted lines, respectively), in panels (a) and (b) using the bottom scales, along with the mean theoretical value for isotropic conditions (thick solid line). Statistical errors (ε_Θ) denoted by diamonds are shown in panels (a) and (b) plotted using the top scales.

errors of AOA measurements. In Figure 5.6, we have plotted both autocorrelation values of AOAs along with AOA time series for an 8-hour interval. The 3-minute AOA values are represented by stars, and the 15-minute smoothed values by thick solid lines. The empirical statistical errors ($\hat{\varepsilon}_\Theta$) are shown by the arrows between the two horizontal lines in (a) and (c). These empirical values are obtained from the difference between the 3-minute autocorrelation values and the autocorrelation of the smoothed values at $\tau = 0$. The theoretical statistical error (ε_Θ) is obtained using Equation 5.29. Comparing these two values, we see that the theoretical value underestimates the empirical value for both baselines (in this example by $\approx 35\%$). This underestimation could be due either to the geophysical variability of the AOAs (notice the short-period oscillations in Figure 5.6 (b)-(d), with amplitudes $< 0.5°$), or

to inherent limitations involved in the approximations used in deriving Equation 5.29.

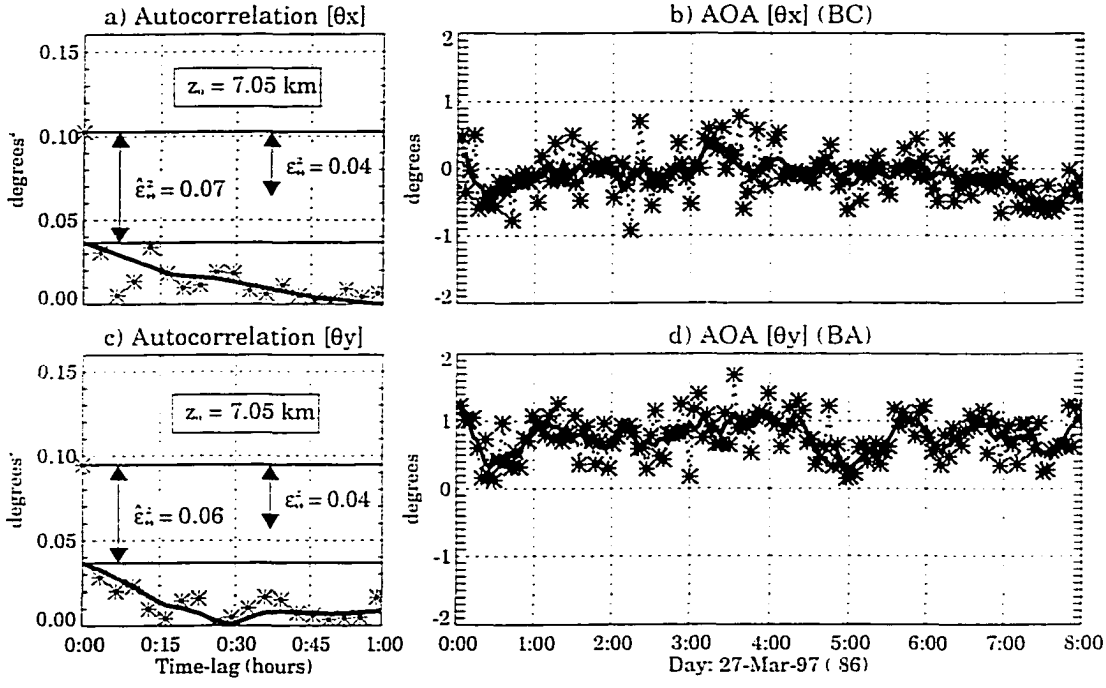


Figure 5.6. Examples of AOA autocorrelation functions ((a) and (c)) and AOA time-series ((b) and (d)) for baselines BC (Θ_x), and BA (Θ_y) respectively. AOA values for the 3-minute measurements are represented in (b) and (d) by stars. Corresponding 15-minute smoothed curves are shown by a thick continuous line. The empirical statistical variances are denoted by $\hat{\varepsilon}^2_\Theta$ (the horizontal lines in (a) and (c)), and the theoretical value by ε^2_Θ (see Equation 5.29). The empirical values are obtained from the difference between the autocorrelation of the 3-minute time-series and the autocorrelation of the smoothed time-series, at $\tau = 0$.

The theoretical and empirical statistical errors translate into CCF phase errors of $\varepsilon_\phi = 7.5°$ and $\hat{\varepsilon}_\phi \approx 9.6°$, respectively. The empirical values are closer to the model results presented by Hocking et al. [1989] (between 8° and 11° for similar coherence and time correlation values). This result, along with the fact that both statistical errors decrease in the same proportion if we use a larger number of points ($M > 768$ in Equation 5.29), results not presented here, suggest that underestimation errors indeed arise from limitations in Equation 5.29. It is possible that a constant of proportionality could be obtained from model simulations like those presented by

May [1988].

In Figure 5.7 we show histograms of correlation times τ_c (panel (a)), obtained from the width of the average auto-correlation function. The coherence values, coh_{BC} and coh_{BA}, for the two orthogonal baselines are shown in panels (b) and (c). The mean values in all panels are represented by dashed vertical lines, and theoretical values under isotropic conditions represented by solid vertical lines. Any difference between these two values are presumably due to anisotropic scattering, i.e., $\xi_h^2 \neq 2a_h^{-2}$.

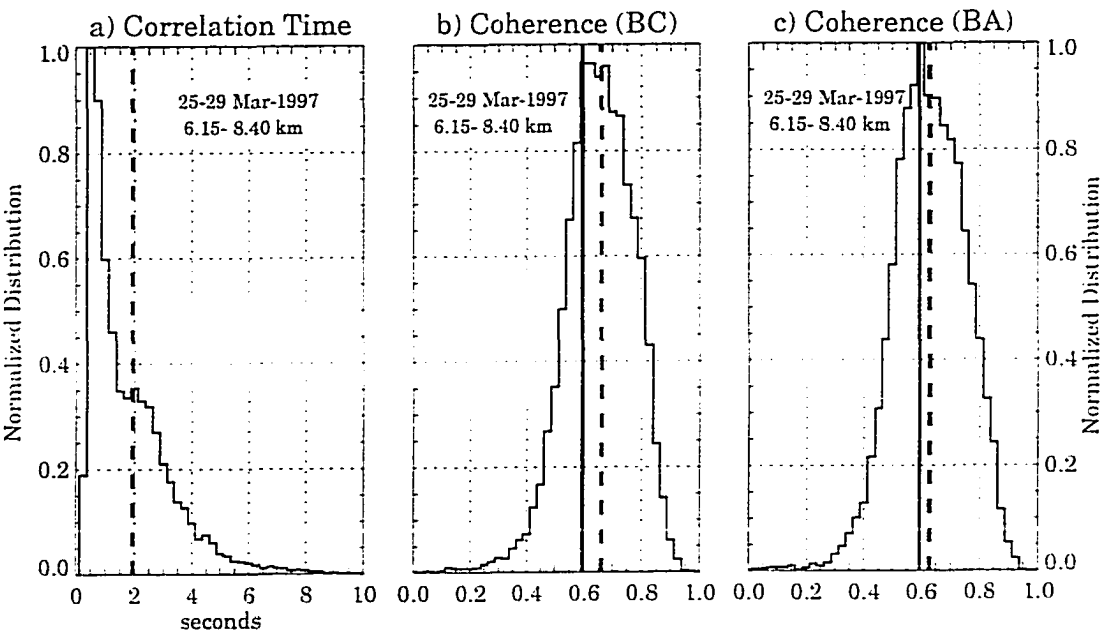


Figure 5.7. Histograms of (a) correlation times (τ_c), and coherence for baselines (b) BC ($\mathrm{coh}_{BC}[0]$) and (c) BA ($\mathrm{coh}_{BA}[0]$). The mean values in all panels are represented with vertical dashed lines, and the theoretical values under isotropic conditions ($\xi_h^2 \approx 2a_h^{-2}$), panels (b) and (c), by vertical solid lines. The anlysis is done for heights between 6.15 to 8.40 km.

In Figure 5.8, we show an $\approx$ 3-day data set of 3-minute AOAs for six separate heights. The theoretical "isotropic" mean values for both baselines, BC and BA, are represented by long dashed horizontal lines. Short dashed lines (in (b)) show the zero values at each height (recall the nonzero value of the BA AOAs, due to the geometry and antenna pointing). Examination of Figure 5.8, we see

the presence of long-period wave-like oscillations (> 12 hours) with amplitudes of $\approx 1°$ (with respect to the theoretical mean values). These oscillations, which exhibit some coherency with height, are related clearly to geophysical causes. Röttger et al. [1990a] also related the AOA periodicities to long-period waves, with amplitudes of 1°-2°. Palmer et al. [1997b] and Tsuda et al. [1997a] have both observed wave-like structure in the azimuth scans of echo power, and relate them to gravity waves effects. These type of AOA oscillations have also been seen in the mesosphere by Meek and Manson [1992], using and MF radar and attributed by them to gravity waves.

5.5 Conclusions

Based on scattering theory, we have identified four separate contributions to AOA lower-atmospheric VHF radar measurements. Specifically, we can isolate contributions from: (a) tilted layers, (b) antenna pointing, and (c) two additional geometrical contributions. We have also shown that the estimation of most AOA contributions requires the measurement of the scattering parameters, under both horizontally isotropic and horizontally anisotropic scattering.

Based upon our theoretical derivation, it appears possible to relate AOA measurements to vertical wind velocities (see Equation 5.20). The correction of vertical velocities can be done without explicitly determining most AOA contributions (e.g. the tilt of reflecting structures).

Measurements of AOAs by non-collocated SA systems have been analyzed and discussed. We have shown that, in theory, vertical velocity corrections can be done with such SA systems. The unambiguous determination of the tilt of reflecting structures depend on the precision of: (a) the SA system (e.g. beam pointing, antenna positions) and (b) the measurement of scattering parameters. Nonetheless, this is also the case for collocated SA systems. The experimental corrections of

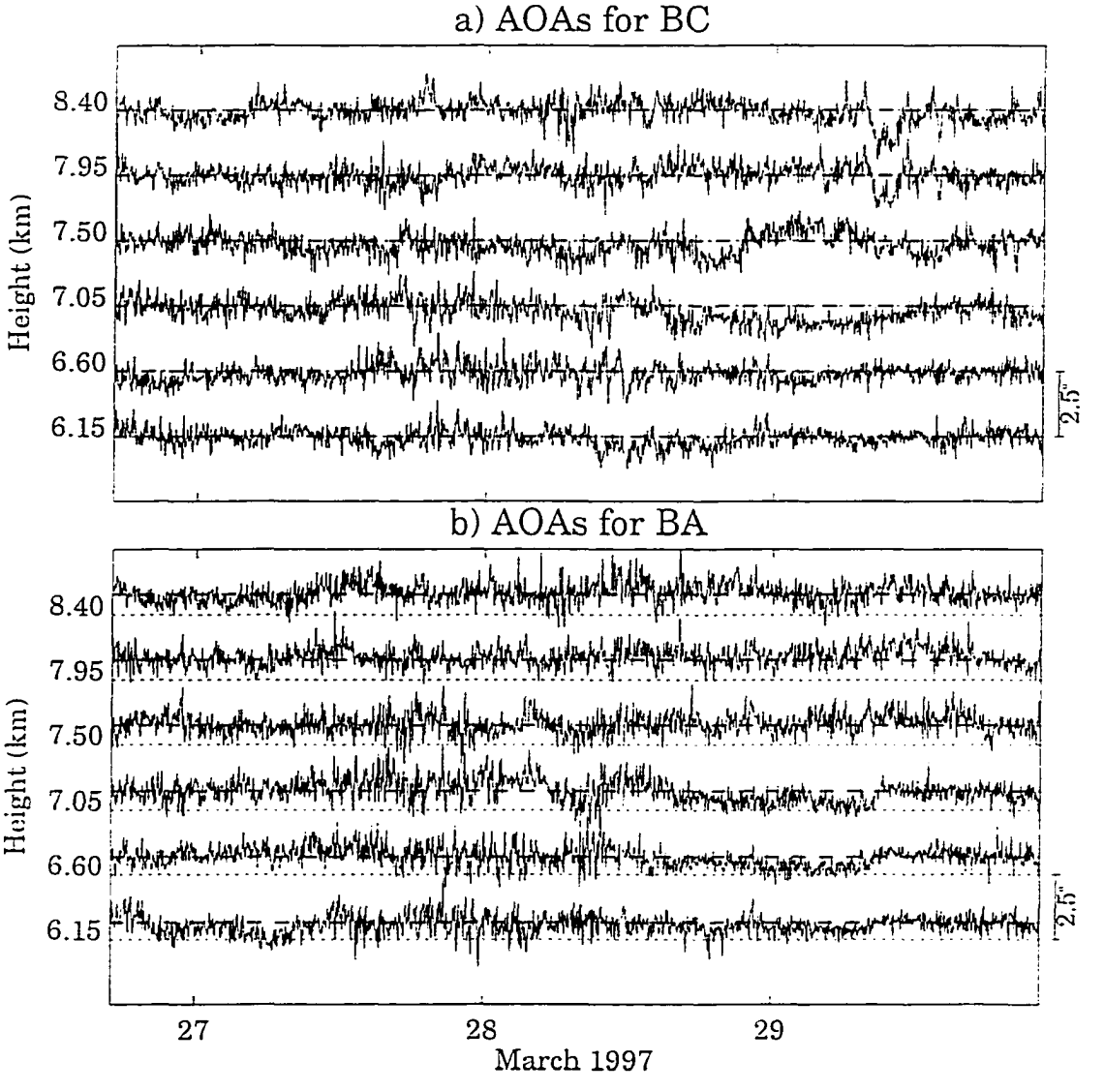


Figure 5.8. Time series of 3-minute AOAs for six tropospheric heights (6.15 to 8.40 km). Results for baseline (a) BC (Θ_x), and (b) BA (Θ_y). Mean theoretical values under isotropic conditions ($\xi_h^2 \approx 2a_h^{-2}$) represented by long dashed lines. The short-dashed lines in panel (b) represent the zero values at each height. Notice the long-period oscillations (> 12 hours).

vertical velocities are currently under study and will be presented in a future work.

We have compared frequency-domain AOA estimates to those from the time-domain method. Both methods give essentially the same information. However, the frequency-domain AOA estimates exhibited more variance (more "outliers") in our implementation, and their accuracy depends on the precision of the Doppler frequency measurement. In addition, the time-domain method is easier to implement, faster to calculate, and more reliable, than the frequency-domain method.

The non-atmospheric AOA contributions, assuming isotropic scattering ($\rho_{ch} \ll a_h^{-1}$), compared reasonably well with the experimental results, particularly below 15 km. Small differences above 15 km are probably due to anisotropic scattering and/or tilted layers.

Given the reasonable agreement between the empirical and theoretical statistical errors discussed here, a strong geophysical component can be seen in our AOA results. This geophysical component exhibit both short- (< 1 hour) and long-period (> 12 hours) oscillations with $\approx 1°$ amplitudes. The detailed characterization of these AOAs measurements and their relationship to gravity waves, vertical velocities and aspect sensitivity will be left for future work.

CHAPTER 6

A STATISTICAL COMPARISON OF VHF TECHNIQUES TO STUDY CLEAR-AIR VERTICAL VELOCITIES IN THE LOWER ATMOSPHERE USING THE JICAMARCA RADAR

6.1 Introduction

Although they are of much smaller magnitude than horizontal motions of the atmosphere, vertical atmospheric motions are very important. Wind profilers offer the unique capability of directly measuring this vertical motion. Such measurements have proven valuable, for example, in defining the character and variability of the spectrum of gravity wave and tidal motions (VanZandt et al. [1991]; Gage et al. [1992]; Liziola and Balsley [1997]). Unfortunately, vertical velocity measurements, particularly mean vertical motions, have proven to be among the most challenging and least reliable of wind profiler measurements. For example, Balsley and Riddle [1984] found a typical value of the mean vertical wind in the summer mesosphere of ≈ -0.3 m s^{-1} that differed significantly from ≈ 0.05 m s^{-1} motion required to account for the summer mesopause thermal structure (e.g. Fritts and Luo [1995]). Although similar results have been reported by others, this discrepancy is not yet understood. Similar discrepancies between expected and observed vertical velocities have been reported by Nastrom and VanZandt [1994] (unreasonably strong downward motions in the tropospheric vertical velocity measurements over Flatland, Illinois), and Fukao et al. [1991] (downward velocities in the troposphere that appear to be coupled with an upward velocity in the stratosphere, in the vicinity of a strong jet stream over Japan).

All of the vertical velocity discrepancies mentioned above were obtained

using VHF radars that employ the Doppler beam-swinging (DBS) technique. The DBS technique uses a single vertically directed beam to measure the vertical velocity. Many of the DBS measurement biases have been attributed to an off-vertical "effective" pointing direction of the vertical beam, which adds a horizontal velocity contamination, or leakage, to the vertical data.

An alternative technique that could take care of this effective pointing problem is the spaced antenna (SA) technique. The SA technique was originally developed to look at ionospheric returns (Briggs et al. [1950]). The technique uses a vertically-directed antenna for transmission and three or more horizontally spaced antennas that are typically vertically directed, for reception (Larsen and Röttger [1989]). With the SA configuration, one can measure the angle-of-arrival (AOA) of the incoming atmospheric signals by measuring the phase difference of complex signals received at different antennas. A corrected vertical velocity could be obtained (Röttger and Ierkic [1985]) by additionally measuring both horizontal and radial velocities. Horizontal velocities are usually calculated from a full correlation analysis (Briggs [1984]), while radial velocities are obtained from the Doppler shift of the incoming signals.

Using this approach, Larsen and Röttger [1991] and Palmer et al. [1991] attempted to explain some of the vertical velocity discrepancies by associating the non-zero AOAs with aspect sensitive tilted layers of refractivity. Aspect sensitivity has been shown to be important at VHF (Tsuda et al. [1986, 1997a, 1997b]) but not at UHF. However, vertical velocity comparisons of concurrent VHF and UHF measurements presented by McAfee et al. [1995] show good agreement, a result that brings into question the aspect sensitivity issue.

To explain the observed large downward vertical velocities over Flatland (Illinois), Nastrom and VanZandt [1994] proposed that the dominant cause of the "apparent" mean vertical motions arises from small changes in the refractive index

induced by gravity waves with a vertical component of propagation. Briefly, for waves with upward energy propagation, the vertical velocity and static stability are negatively correlated: therefore, in such waves, regions of downward moving air will have a higher radar reflectivity on a statistical basis and will produce apparent downward velocities. This hypothesis has been supported by the mesospheric studies of Hoppe and Fritts [1995].

Palmer et al. [1997a] and Larsen and Palmer [1997] have suggested that gradients in the wind field can also affect vertical velocity estimates, particularly in radars that have relatively large beam-widths. This effect could explain intriguing results like those presented by Fukao et al. [1991].

The theoretical studies presented by Muschinski [1996] and Palmer et al. [1997a], show that measurements done with a very narrow antenna beam should be less contaminated by both aspect sensitivity and wind-field divergence effects.

In order to better assess the problems associated with VHF radar vertical velocity measurements, we have compared a number of the above techniques using a common data set. We have used the VHF radar at Jicamarca, Perú for these measurements. The Jicamarca antenna can be configured for a number of simultaneous measurements. The beam-width using the entire antenna, moreover, provides the narrowest VHF beam of any of the atmospheric radars. Taking these facts into account, we have made vertical velocity measurements with the narrow beam and used it as our reference profile in the current vertical velocity comparison studies. This is a reasonable approach, particularly since the validation of vertical velocity measurements is strongly hampered by the lack of a standard for comparison (there is no standard measurement of vertical velocity). In addition to the reference profile, we concurrently measured vertical velocities using (a) a wide beam direct measurement, (b) a "quasi" DBS inferred measurement, and (c) four corrected vertical measurements using the "classical" SA approach (these terms will be defined in the following

sections).

A description of the various vertical velocity techniques is given in Section 6.2. Details of the Jicamarca experiment follows in Section 6.3. The results of the vertical velocity comparisons are presented in Section 6.4. Finally, a discussion and a summary of our results are provided in Section 6.5 and Section 6.6, respectively.

6.2 VHF radar techniques for vertical wind velocity measurements

In this section, we describe two of the main techniques (Zenith and spaced antenna techniques) currently used by VHF radars for measurements of the vertical wind velocity (w). Sub-groups of these techniques are divided into separate sections for convenience.

6.2.1 The Zenith technique

Zenith techniques use a single vertically-directed antenna to directly measure w. Such systems operate in a monostatic mode, i.e., the same antenna is used for transmission and reception, to directly measure w. This technique corresponds to the vertical portion of a conventional DBS method.

One of the main disadvantages of the Zenith technique lies in the accuracy of the antenna pointing direction. Small deviations from vertical can caused significant horizontal velocity contributions on the measured vertical velocities (Huaman and Balsley [1996]), particularly in the mean vertical velocity values. Pointing errors in vertically-directed antennas can be caused by systematic problems and/or atmospheric contributions. Pointing errors due to systematic problems are usually small, and ways of measuring and correcting them have been presented by Balsley et al. [1988] and Huaman and Balsley [1996]. Atmospheric contributions, on the other hand, are usually attributed to "off-vertical" centered aspect sensitivity functions (e.g. Röttger and Ierkic [1985]; Röttger et al. [1990a]; Larsen and Röttger [1991]).

Pointing errors due to these types of atmospheric contributions are expected to be proportional to the antenna beam-width (Muschinski [1996]), i.e., the narrower the beam the smaller the deviation from vertical.

In this chapter we compare two different Zenith technique measurements using: (a) a narrow beam and (b) a wide beam antenna. The two-way half-power beam-widths (HPBW) are $\approx 0.85°$ and $\approx 3°$ for the narrow and wide beam antennas, respectively. Hereafter, narrow beam results are denoted by Wn, and wide beam results by Ww. Both measurements were analyzed in the frequency-domain using NOAA's algorithm for spectral moment estimation described in Appendix B.

6.2.2 Spaced-antenna (SA) techniques

In contrast to the Zenith technique, SA techniques provide inferred measurements of w. SA techniques rely on radial velocities made in different pointing directions and make use of the following expression.

$$v_r \approx u \sin\Theta_u + v \sin\Theta_v + w(1 - \sin^2\Theta_u - \sin^2\Theta_v)^{1/2} \qquad (6.1)$$

where v_r is the radial velocity (positive going away from the radar). Θ_u and Θ_v are the zonal and meridional beam direction angles, and where zonal, meridional and vertical wind velocities are denoted by u, v and w, respectively.

From Equation 6.1, SA techniques are subdivided into "quasi" DBS and "classical" SA techniques. These techniques are outlined separately below.

6.2.2.1 The "quasi" DBS approach

The "quasi" DBS approach requires at least three different beam positions. The three-dimensional wind vector is obtained by solving a set of at least three equations like Equation 6.1, where the radial velocities and the pointing directions are known. The main assumptions in this approach are: (a) u, v and w remain the same during the integration time and throughout the volume defined by the different

beams, and (b) the pointing directions Θ_u and Θ_v are well known.

We have implemented this approach at Jicamarca by synthesizing nine different beams through software (see Appendix D for details). This implementation is usually termed post statistics steering (PSS) and was introduced by Kudeki and Woodman [1990]. Once the beams are synthesized, the radial velocities can be calculated via NOAA's algorithm (Appendix B) used in the Zenith techniques. The beam directions are calculated considering the geometry involved, the height sampled, the antenna patterns and assuming an isotropic atmosphere. A similar approach has been used by Palmer et al. [1993], but using just three beam positions. The use of nine beams allows the possibility of averaging out the errors introduced by anisotropic scattering.

6.2.2.2 The "classical" SA approach

The "classical" SA approach uses only one beam direction (usually vertical) and multiple antennas (Röttger and Ierkic [1985]). This technique requires knowledge of the angle-of-arrivals (effective pointing directions) and the radial and horizontal velocities. The vertical velocity, also known as "corrected" vertical velocity (e.g. Larsen and Röttger [1991]), is obtained from

$$\hat{w} = v_r - (u \sin \Theta_u + v \sin \Theta_v) \tag{6.2}$$

where we are assuming small angles. The horizontal "leakage" on the radial velocity, due to "off-vertical" AOA contributions (atmospheric and systematic) can be removed to get a "corrected" w measurement ($\hat{w}$).

In the present study we have calculated the effective beam positions (Θ_u, Θ_v), also called angle-of-arrivals (AOAs), by calculating the phase differences between receiving antennas. To improve reliability, we have made these calculations in the time-domain (see Chapter 5 for details). In addition, we have also determined radial velocities (v_r) in the time-domain by measuring the phase slope at zero-lag of

the complex auto correlation functions (ACFs) (e.g. Woodman [1985]). The horizontal velocity measurements have been obtained from four different SA techniques (see Chapter 4) in order to study their effects on vertical velocity corrections. The four techniques include three separate "true" velocity estimates using full correlation analysis (FCA), and a "true" velocity estimate using a full spectral analysis (FSA). The three FCA "true" velocity estimates are obtained using techniques reported by Briggs [1984] (here: FCAB), Meek [1980] (here: FCAM), and Holloway et al. [1997a] (here: FCAH). The "corrected" w obtained by using the FSA horizontal velocity (here: FSAM), on the other hand, is a modification to the approach presented by Briggs and Vincent [1992]. We obtain the horizontal velocity from a frequency-domain analysis. The radial velocity (v_r) and the AOAs are obtained from a time-domain analysis.

Salient characteristics of the aforementioned SA techniques for horizontal wind measurements are summarized in Table 6.1. Additional details of these techniques can be found in Chapter 4, where the FCAMt and FCAH(r) there, corresponds to FCAM and FCAH in this chapter.

Table 6.1. Salient features of the SA techniques for horizontal wind measurements used in vertical wind velocity estimation.

Technique	Reference	Scattering[a]	Turbulence	Domain
FCAB	Briggs [1984]	Anisotropic	Yes	Time
FCAM	Meek [1980]	Anisotropic	Yes	Time
FCAH	Holloway et al. [1997a]	Isotropic	Yes	Time
FSAM	Briggs and Vincent [1992]	Anisotropic	Yes	Frequency

[a]Type of scattering in the horizontal plane

6.3 Experimental setup

The Jicamarca radar (near Lima, Perú) used for these studies operates at $\approx$ 50 MHz and has a 288 m by 288 m antenna array. The array is composed of 64 separate modules. Each module consists of a 12 by 12 array of cross-polarized

half-wave dipoles. Figure 6.1 shows the antenna configuration for our measurements. The vertically-pointing full array (narrow beam, E) was used for both transmission and reception in the "down" polarization (i.e., the lower set of dipoles [x aligned]). The "up" polarization (the orthogonal upper dipole set [y aligned]) was used for the vertically-pointing D antenna (wide beam) for both transmission and reception. The on-axis-directed A, B and C modules were used for reception only in the "up" polarization. Note that the on-axis pointing direction is not vertical but rather $\approx -1.46°$ from vertical along the the y-axis, i.e. toward the Southwest and results from the slope of the valley floor. The characteristics of these five antennas have been summarized for convenience in Table 6.2.

Table 6.2: Antennas characteristics.

	Antennas				
	A	*B*	*C*	*D*	*E*
Area (m^2)	36x36	36x36	36x36	72x72	288x288
No. Modules	1	1	1	4	64
One-way HPBW	8.18°	8.18°	8.18°	4.05°	1.15°
Direction	On-axis	On-axis	On-axis	Vertical	Vertical
Transmission	No	No	No	Yes	Yes
Reception	Yes	Yes	Yes	Yes	Yes
Polarization	"up"	"up"	"up"	"up"	"down"

The outputs of each of the five receiving antennas were independently received, sampled, and recorded. Ground clutter effects were partially filtered separately in both the real and imaginary signal channels.

We obtained a 5-day data set in March, 1997 (from March 25, 16:50 LT to March 29, 22:15 LT) running simultaneously in the SA and Narrow-Wide beam modes. Operational parameters for this data set are presented in Table 6.3. Due to availability of only four receivers at the time of the experiment, complex signals from antennas D and E shared the same physical receiver. This process was possible using two sampling windows with an effective delay of 1 ms. After digitization, the two complex signals were recovered and recorded in independent channels (via software).

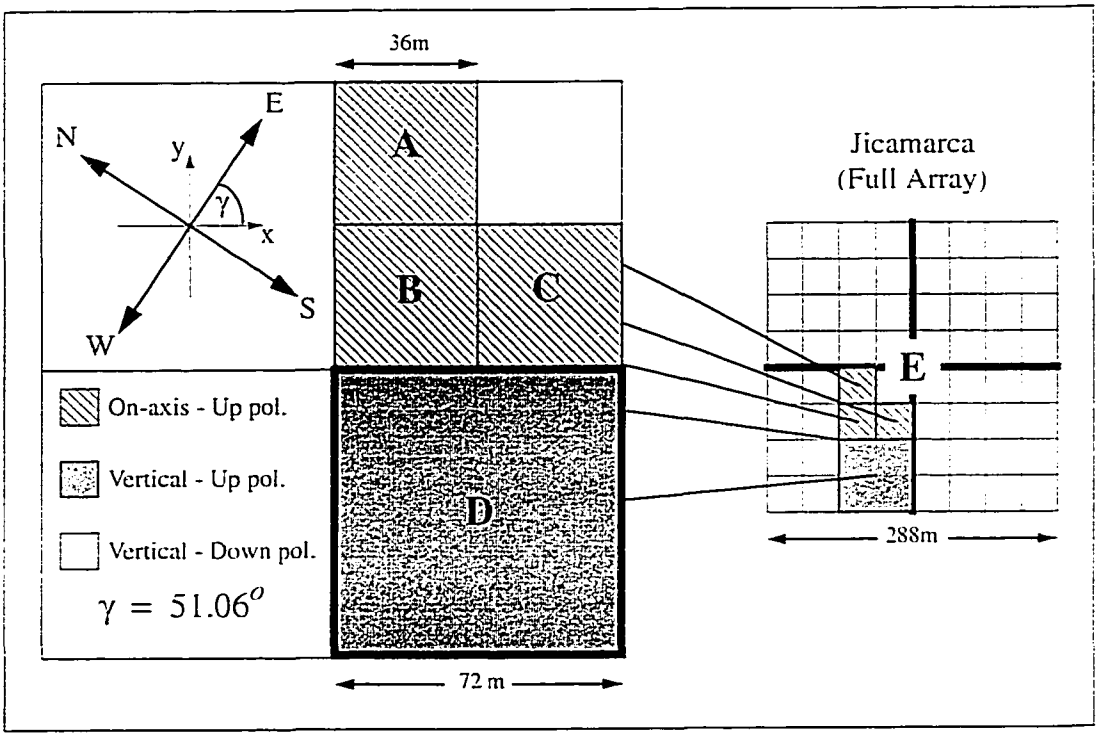


Figure 6.1. Antenna setup for the Jicamarca experiment. The full array, vertically pointing E module, was used for transmission and reception in the "Down" polarization (x-aligned). In the orthogonal polarization ("Up", y-aligned), the vertically pointing D module was used for transmission and reception. The "on-axis" pointing A,B and C modules are used only for reception. The "on-axis" position differ from the zenith position, because the antenna surface is not exactly horizontal, but is tilted $\approx 1.46°$, approximately towards the Southwest. Notice that neither of the antenna axes, (x, y), is aligned with the geographic coordinates (E, N).

Independent measurements were obtained every ≈ 3 minutes during this period for all heights between 3 km and 21 km for each antenna/receiver set. The resulting data set was used to analyze all of the techniques studied.

A calibration procedure was performed before and after each data set in order to determine the phase differences between the receiving lines (more details can be found in Appendix C). The ground clutter removal and calibration processes, were done in the same way described in Chapter 4.

Table 6.3. Operating parameters for the experiments conducted at Jicamarca (from 25-Mar-97 to 29-Mar-97).

Parameter	Value
Pulse repetition period	2 ms
Number of coherent integrations	128
Number of incoherent integrations	6
Number of points	128
Effective sampling time	256 ms
Height resolution	0.45 km
Initial height	3 km
Number of heights	40
Transmitting power at antenna D	100 kW
Transmitting power at antenna E	30 kW

6.4 Results

As mentioned in the Section 6.1, we adopted a "reference profile" in order to quantify the agreement of the different w measurements. We elected to use the narrow beam measurements (Wn) as our reference. These narrow beam measurements were expected to have less vertical velocity biases than those obtained using a wide beam (see Section 6.2). In addition, Wn measurements had a higher signal-to-noise ratio (SNR) owing to the large antenna area.

All of the vertical velocity measurements (both direct and inferred) were "cleaned" if the SNR was low (< -6 dB). In addition, inferred vertical velocities were edited out when either the zonal or meridional velocities were cleaned. Horizontal velocities were cleaned twice (pre-cleaning and post-cleaning) in the same manner as that described in Chapter 4.

Statistical comparisons were made over three different time periods: (1) the entire 5-day data set (All), (2) a selected 24-hour Quiet period, and (3) a selected 24-hour Active period. Additional information on these periods has been tabulated in Table 6.4.

Before presenting the results of our vertical velocity analyses, it will be useful to provide selected background results for these selected periods:

Table 6.4. Information on the periods (All, Quiet and Active) used in these analyses.

Parameter	Data set		
	All	Quiet	Active
Starting date	25-Mar-97	26-Mar-97	28-Mar-97
Starting time	17:00	16:00	22:00
Duration	5 days	24 hours	24 hours
Conditions	quiet/ active	quiet	active

• SNR profiles. SNR profiles are shown in Figure 6.2 for each period (All, Quiet and Active). A $\overline{\text{SNR}}$ profile for the SA configuration (the average SNR for antennas A, B and C) is plotted with pluses; profiles using the wide and narrow beam vertical antennas are shown by asterisks and diamonds, respectively. Clearly the largest SNR is recorded using the narrow beam antenna, particularly above 10 km. This statement is true even though the transmitted power on this antenna was ≈ 5 dB less than the transmitted power on the wide beam antenna. Below 10 km the SNR between the wide and narrow beams are more comparable. This is primarily due to the fact that this region lies in the near-field of the narrow beam antenna. Comparison of panels (b) and (c) in this figure suggests an enhancement of SNR during the active period relative to the quiet period (compare, for example, the SNRs at 7.05 km, where the active value is > 10 dB relative to the quiet value). Above 15 km, the peak $\overline{\text{SNR}}$ enhancement in the three panels suggests that the tropopause height is ≈ 17 km (Riddle et al. [1984]).

Based on these SNR profiles, hereafter we will show results only for heights with SNR > -6 dB. Also, we will exclude in our analysis heights below 5 km, where problems due to receiver recovery are known to exist.

• Turbulence intensity and correlation length. Using a full correlation analysis (FCA) we were able to obtain estimates of the turbulence intensity (σ_t) and the correlation length (ξ'). Turbulence intensity was determined from,

$$\sigma_t = \left[\frac{1}{2k_0^2}\left(\frac{1}{2\tau_c^2} + 4Av_{0x}^2 + 4Bv_{0y}^2 + 8Hv_{0x}v_{0y}\right)\right]^{1/2} \tag{6.3}$$

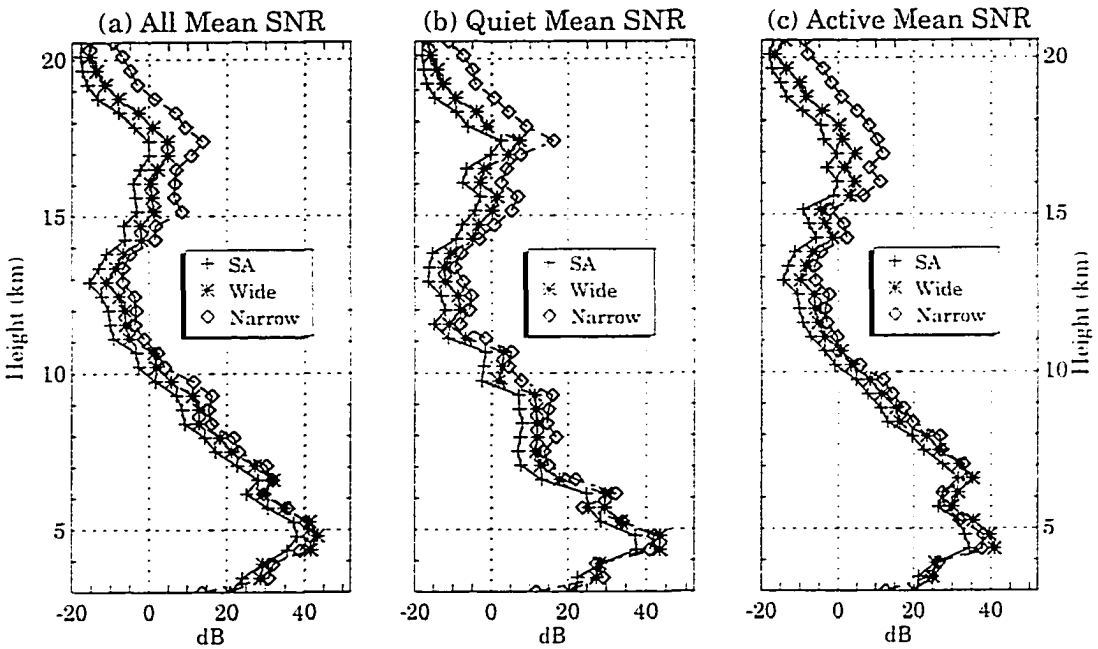


Figure 6.2. Mean SNR profiles for the following time periods: (a) All (b) Quiet and (c) Active. The SA $\overline{SNR}$ profile is denoted by "+s", while the wide beam and narrow beam $\overline{SNR}$ are denoted with asterisks ($*$) and diamonds ($\diamond$) respectively.

where τ_c is the correlation time, k_0 is the radar wavenumber, (A,B H) are FCA parameters, and v_{0x} and v_{0y} are the velocity projections in the x and y axes (see Holloway et al. [1997b] for more details).

The correlation length ξ' represents the average value of the major and minor axes of a horizontal aspect sensitivity function, and is equal to $(\xi'_x + \xi'_y)/2$ in Holloway et al. [1997b] and $\propto (a+b)/2$ in Meek [1980]. Thus, the bigger ξ', the more aspect sensitive the atmosphere.

Median values of these two parameters are plotted in Figure 6.3, for the three time periods. Examination of Figure 6.3 below 10 km suggests that the active period is considerably more turbulent and somewhat less aspect sensitive than the quiet period. It is also apparent that for both active and quiet periods, the region above 15 km appears to be more aspect sensitive than the lower heights.

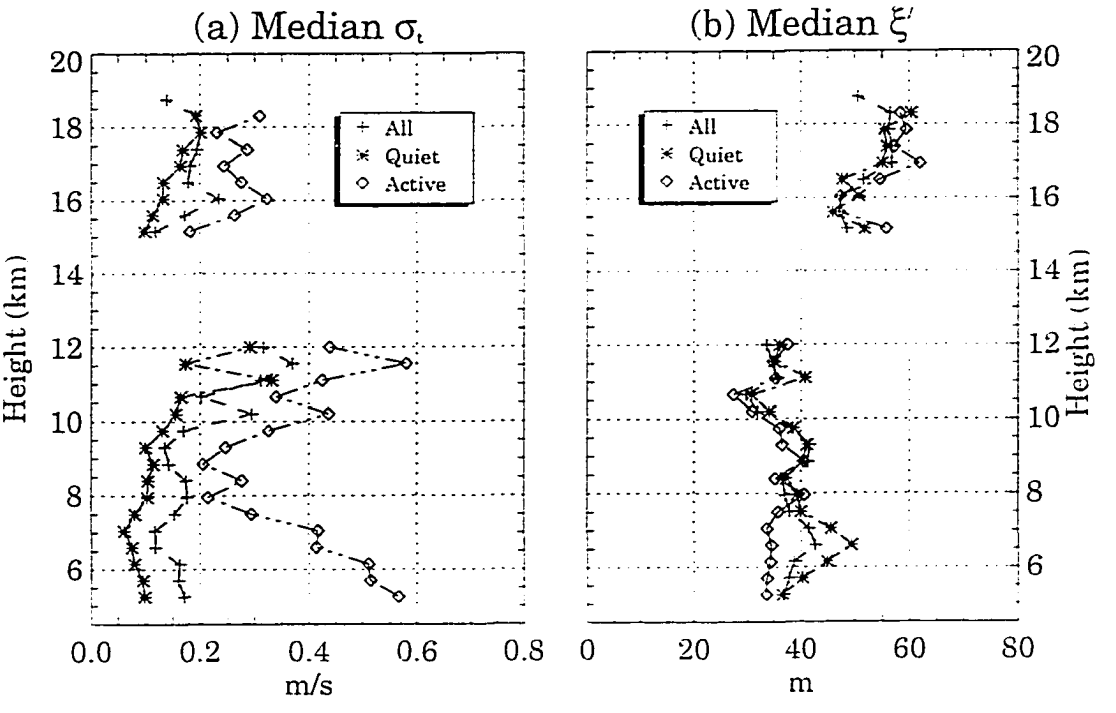


Figure 6.3. (a) Median turbulence intensity (σ_t), and (b) median correlation length (ξ'), for the time-periods: All (+), Quiet (*) and Active ($\diamond$).

• Horizontal wind velocity profiles. We have calculated zonal and meridional velocities from the SA configuration by applying some of the techniques used in Chapter 4. In Figure 6.4, the zonal (upper panels) and the meridional (lower panels) wind components are plotted for each of the three time periods. Examination of Figure 6.4 shows that, over all, there is a stronger meridional component during the active period, than during the quiet period. Furthermore, at near-tropopause heights (i.e., above $\approx$ 15 km), the zonal wind is westward and somewhat stronger during the active period ($u \approx -20$ m s^{-1}) than during the quiet period ($u \approx -10$ m s^{-1}).

In general the agreement among the SA techniques can be summarized as: good below 12 km and poor above 15 km. The discrepancy above 15 km is particularly obvious in the zonal component. At these heights, FCAB (with asterisks) and FCAH (with diamonds) measurements are somewhat larger than the FSAM (with triangles) and PSS (with squares) measurements. A detail discussion on the horizontal velocity

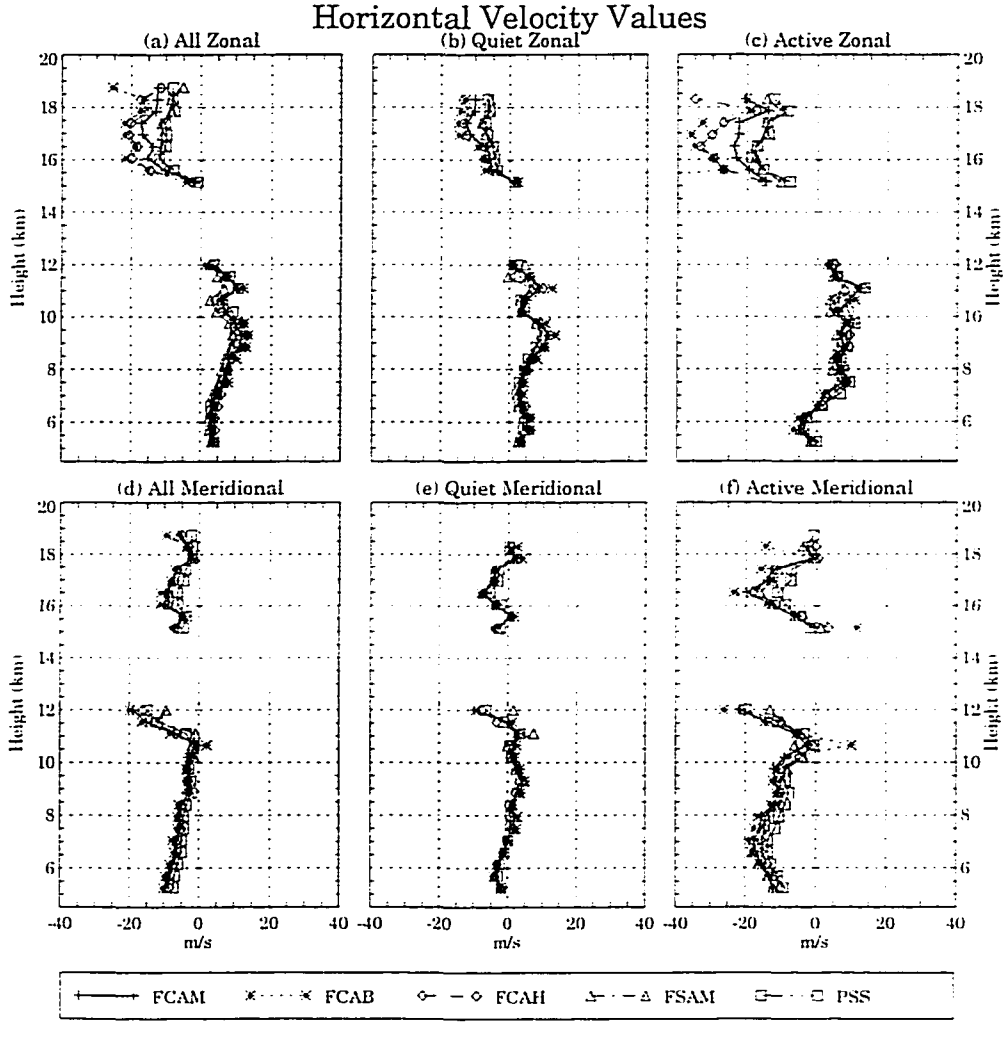


Figure 6.4. Profiles of mean zonal and meridional wind velocities (μ_u and μ_v) for the following SA techniques: the reference profile FCAM (+), FCAB (∗), FCAH (◇), FSAM (△), and PSS (□). Each column corresponds to a particular selected time period (All: first column, Quiet: second column, and Active: third column).

agreement of these SA techniques is given in Chapter 4.

• AOA profiles. Median values (diamonds) with standard deviations (horizontal solid lines) of observed zonal and meridional AOAs are plotted in Figure 6.5. In addition, we show the AOA (solid curve) values that would be expected for scattering from an isotropic atmosphere, given the antenna geometry and pointing beam directions. Note that these theoretical values do not include the geometrical contributions that are independent of the horizontal correlation scales (Λ) discussed in Chapter 5. These geometrical contributions were removed from the AOAs plotted in Figure 6.5 and from the AOAs used in Equation 6.2 to get "corrected" vertical velocities ($\hat{w}$). AOA results show clearly that the echoes are coming from small "off-vertical" angles. This effect, on the other hand, is reasonably predictable, and clearly arises from geometric, pointing beam direction and atmospheric effects. Geometric and pointing beam direction effects alone would produce AOAs shown by the solid curve; additional atmospheric effects would produce departures from the solid curve. Note that the standard deviations in this figure can be seen to be as high as 0.6°. Given the results shown in Figure 6.5, radial velocities measured with the SA system must be contaminated by the horizontal velocity. Clearly, a correction to these "vertical" velocities is required in order to get "true" vertical velocities. Usually, SA radial velocities are calculated by averaging the radial velocities of all the receiving antennas. This averaging process, however, will still cause biases in the "corrected" vertical velocities if a non-collocated SA system is used. These biases are caused by the geometrical contributions that depend on the horizontal correlation scales (Γ) discussed in Chapter 5. Note that, in a non-collocated SA system, the geometrical center of the array of receiving antennas does not coincide with the geometrical center of the transmitting antenna. Since our SA system at Jicamarca is non-collocated, SA radial velocities (FCAr) that are used in Equation 6.2 to get $\hat{w}$ have been calculated from the average of the radial velocities of antennas A and B in order to avoid the

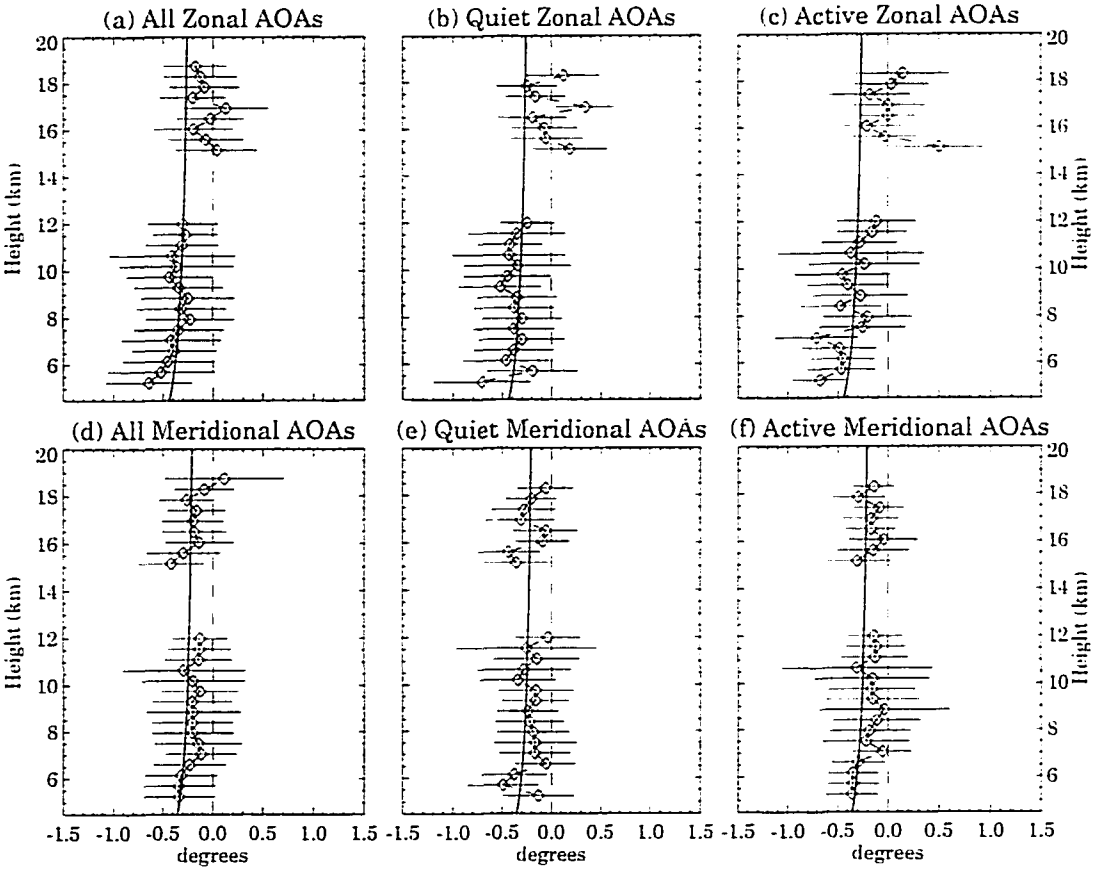


Figure 6.5. Zonal and meridional AOAs for the three selected time-periods: All (first column). Quiet (second column) and Active (third column) time periods. Median values are shown by diamonds and standard deviations are shown by horizontal solid lines. The solid curve in each plot represents theoretical values from Chapter 5 expected from an isotropic atmosphere, given the antenna geometry and pointing direction. The vertical dashed lines represent AOAs of zero degrees.

biases due to the variable geometrical contributions.

• Some vertical velocity characteristics. In Figure 6.6 we provide pertinent results for the standard deviation (σ_w) and percentage of failures (α_w) of the vertical velocity profiles for all three periods. Results are shown for the following techniques: a narrow beam Zenith technique (Wn), a "corrected" vertical velocity technique (FCAM), a "quasi" DBS technique (PSS) and a wide beam Zenith technique (Ww).

Examination of Figure 6.6 shows large variations of the percentage of failures with

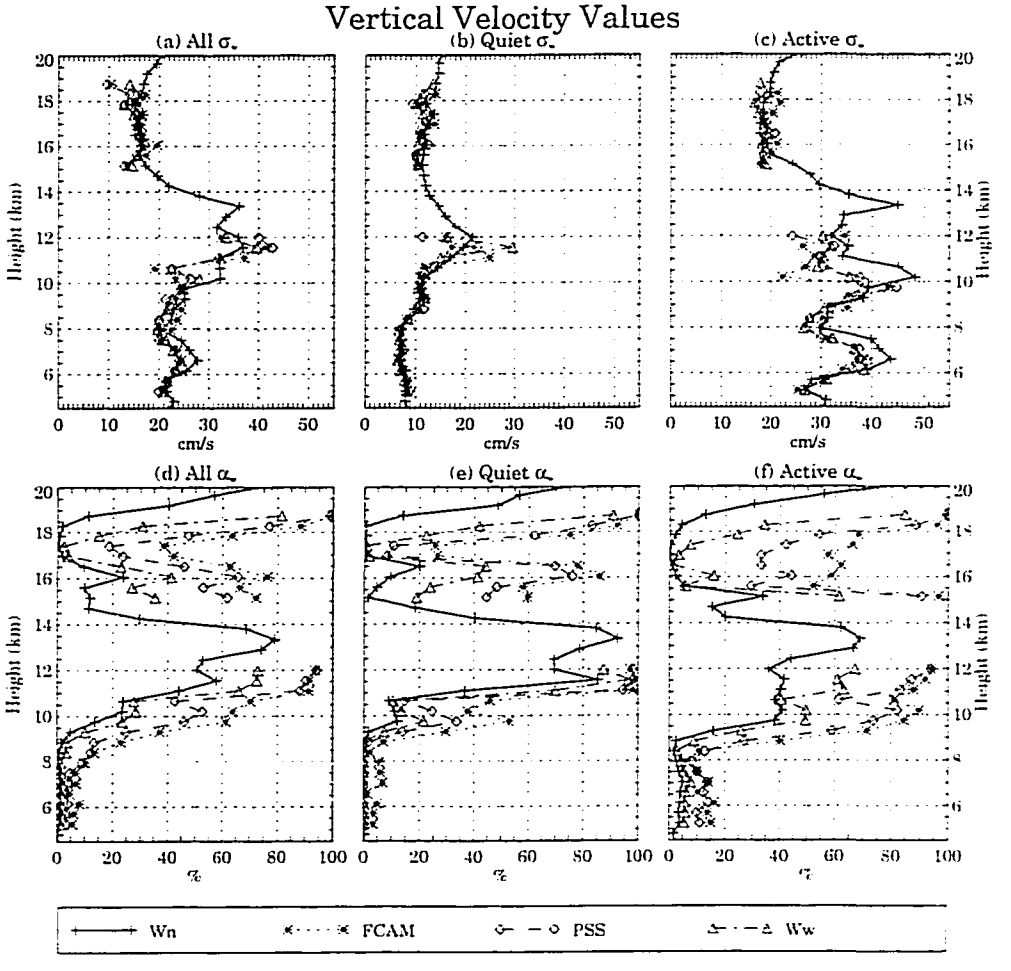


Figure 6.6. Profiles of vertical velocity characteristics: standard deviations (σ_w, upper panels) and percentage of failures (α_w, lower panels), for All (first column), Quiet (second column) and Active (third column) time periods. The w measurements are: Wn (+), FCAM (*), PSS ($\diamond$) and Ww ($\triangle$).

heights, with the largest failures occurring in the regions of weakest SNR (see Figure 6.2). Also, the inferred measurements (FCAM, PSS) exhibit large values of α ($> 10\%$) below 12 km and during the active conditions. The standard deviation estimates of all the techniques show $\sigma_w < 20$ cm s^{-1} during the quiet period, and $\sigma_w \geq 20$ cm s^{-1} during the active period. However, during the active period and below 12 km, the variance of the narrow beam technique is larger than the variance of the other techniques.

We present in Figures 6.7 and 6.8 a statistical comparison of various vertical velocity techniques. Using Wn as a reference, we first compare four "classical" SA techniques (FCA-type) in Figure 6.7. The other vertical velocity measurements (PSS, Ww and FCAr) are then compared against the same reference in Figure 6.8. All comparisons are done in terms of: (a) mean velocity difference ($\mu_{\Delta w}$), (b) standard deviation of the velocity difference ($\sigma_{\Delta w}$), and (c) correlation coefficient between velocities ($\rho_{\Delta w}$).

Examination of Figure 6.7 shows excellent agreement among the four "corrected" vertical velocity measurements ($\hat{w}$) at tropospheric heights, where all the techniques give similar values of $\mu_{\Delta w}$, $\sigma_{\Delta w}$ and $\rho_{\Delta w}$. Above 15 km, however, some discrepancies can be seen, particularly for the active period. Note that the mean differences ($\mu_{\Delta w}$) of $\hat{w}$ tend to be negative below 12 km and positive above 15 km during all three periods.

The general characteristics of the $\hat{w}$ standard "errors" ($\sigma_{\Delta w}$) apparent in Figure 6.7 are that they: (a) increase significantly with increasing height up to 12 km, (b) are relatively constant and small above 15 km, and (c) increase at all heights with increased wind activity. The $\hat{w}$ correlation coefficients, on the other hand, tend to decrease slightly with increasing height, and increase at all heights with increased wind activity.

Comparable statistical comparisons for the other ("uncorrected") vertical

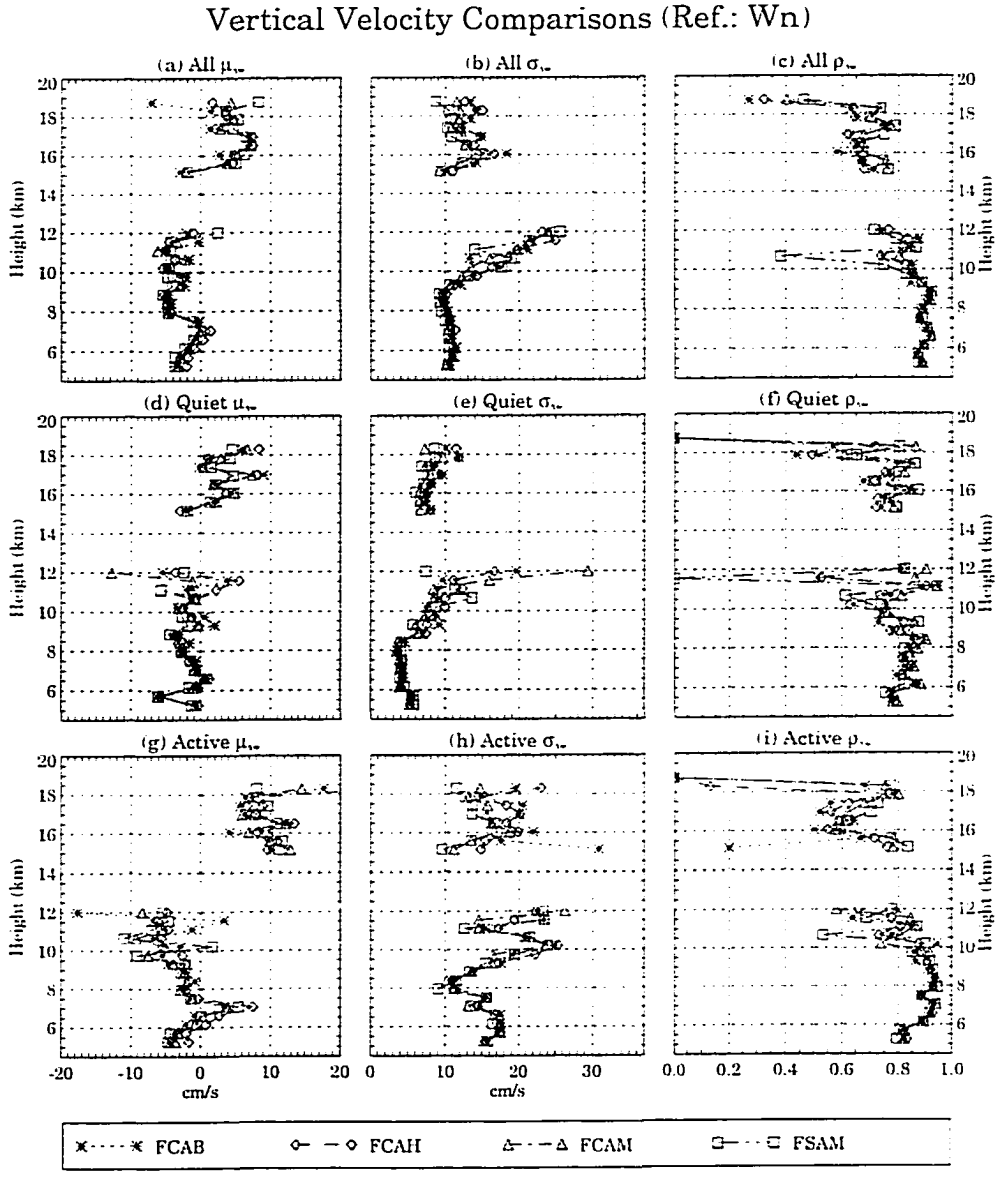


Figure 6.7. Profiles of vertical velocity comparisons for the following FCA-type measurements: FCAB (∗), FCAH (◇), FCAM (△) and FSAM (□), using Wn as a reference. The comparisons are in terms of: mean velocity differences ($\mu_{\Delta w}$, first column), standard deviation of the velocity differences ($\sigma_{\Delta w}$, second column) and correlation coefficient ($\rho_{\Delta w}$, third column), for All (first row), Quiet (second row) and Active (third row).

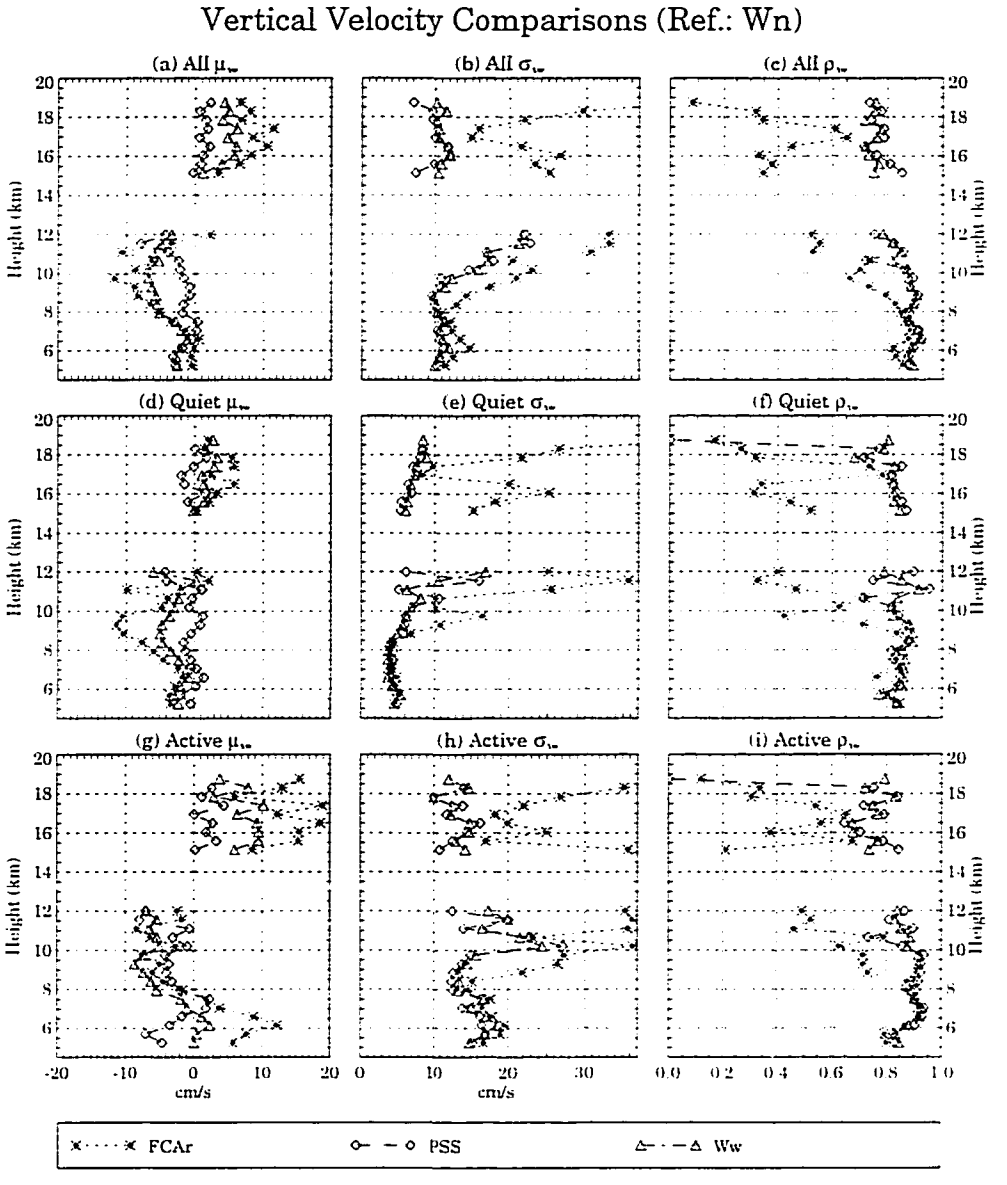


Figure 6.8. Profiles of vertical velocity comparisons for the following measurements: FCAr (∗), PSS (◇) and Ww (△), using Wn as a reference. The comparisons are in terms of: mean velocity differences ($\mu_{\Delta w}$, first column), standard deviation of the velocity differences ($\sigma_{\Delta w}$, second column) and correlation coefficient ($\rho_{\Delta w}$, third column), for All (first row), Quiet (second row) and Active (third row).

velocity measurements appear in Figure 6.8. To facilitate comparisons with Figure 6.7 results, we have included the "off-vertical" radial measurements (FCAr). It is worth noting that the biggest biases and errors and smallest correlations are given by these radial velocities.

The standard "errors" and correlation coefficient results for the Ww (triangles) and PSS (diamonds) techniques are very similar to the "corrected" results presented in Figure 6.7. The mean differences ($\mu_{\Delta w}$), on the other hand, show significant discrepancies: Ww differences are larger and PSS differences smaller than the differences shown in Figure 6.7.

In order to better see these differences, we have plotted in Figure 6.9 the mean values of the two Zenith techniques (Wn and Ww), in addition to a "classical" SA technique (FCAM) and the "quasi" DBS technique (PSS). Examination of Figure 6.9 shows that the Ww measurements tend to give downward velocities below 12 km, and upward velocities above 15 km. Similar comments can be applied to the FCAM results, although the Ww measurements tend to give more biases than the FCAM results, even when Wn is upward, i.e., at mid-tropospheric heights.

On the other hand, PSS results are in somewhat better agreement with the reference, particularly during quiet conditions ($|\mu_{\Delta w}| \leq 1$ cm s^{-1}). During active conditions, the PSS measurements also give negative mean differences at tropospheric heights and positive mean differences above 15 km relative to the reference profile (see Figure 6.8). Nonetheless, these mean differences are, in general, less than those obtained from the other techniques.

In Figure 6.10 we present the standard deviations of the vertical velocity difference (using Wn as reference), and the correlation coefficients between both wide and narrow beam measurements (stars). In addition we present the approximated values of the statistical errors (σ_ϵ) of the "instantaneous" vertical velocities (diamonds in upper panels), and theoretically-derived correlation coefficients ($\hat{\rho}_{\Delta w}$)

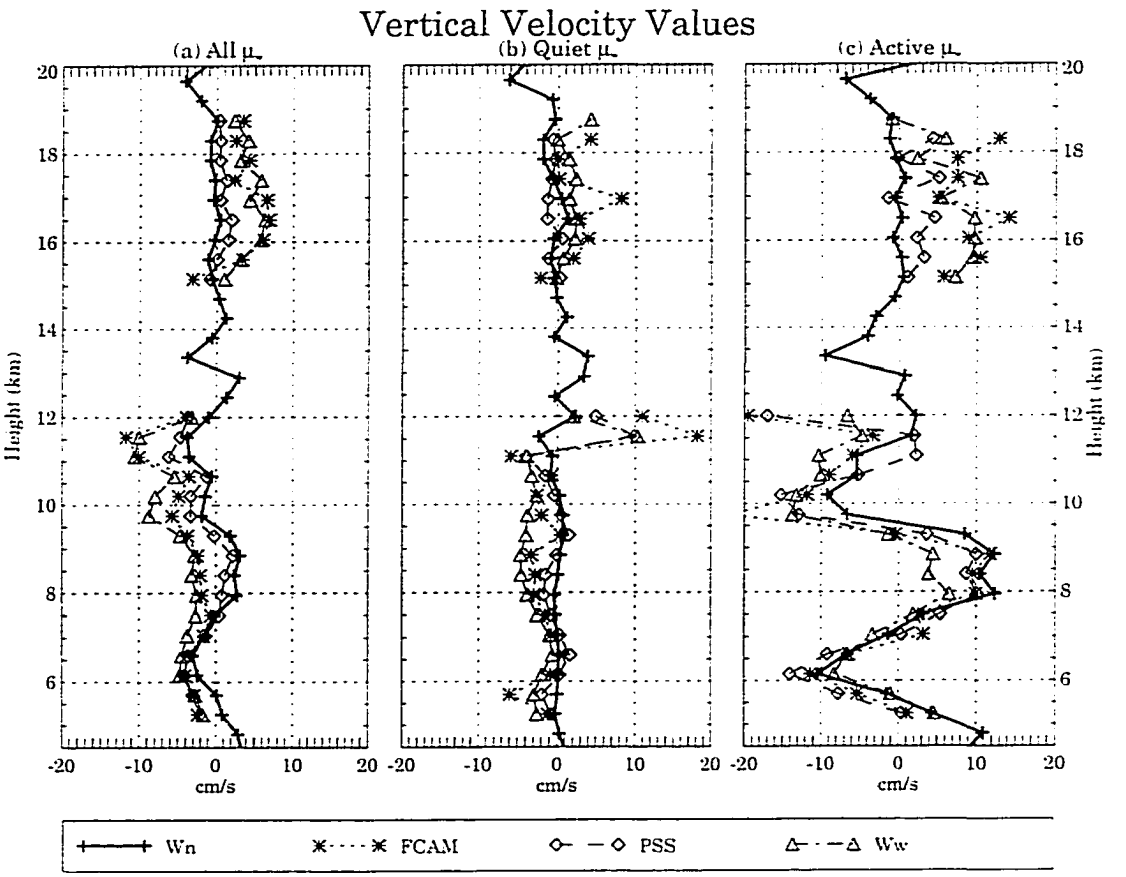


Figure 6.9. Profiles of mean vertical velocities (μ_w) for time-intervals of events (a) All, (b) Quiet and (c) Active. The measurements are: the reference profile Wn(+). FCAM (∗), PSS (◇), and Ww (△).

(diamonds in lower panels). The last two quantities have been obtained from:

$$\sigma_\epsilon^2 = \frac{\sigma_{\Delta w}^2}{2} \tag{6.4}$$

$$\hat{\rho}_{\Delta w} = \frac{\hat{\sigma}_w^2}{\sqrt{\hat{\sigma}_w^4 + \hat{\sigma}_w^2 \sigma_{\Delta w}^2 + \sigma_{\Delta w}^4/4}} \tag{6.5}$$

We have assumed that,

$$W_n = \bar{w} + w + x \tag{6.6}$$

$$W_w = \bar{w} + \Delta w + w + y \tag{6.7}$$

where W_n is the narrow beam and W_w the wide beam vertical velocities. $\bar{w}$ is the

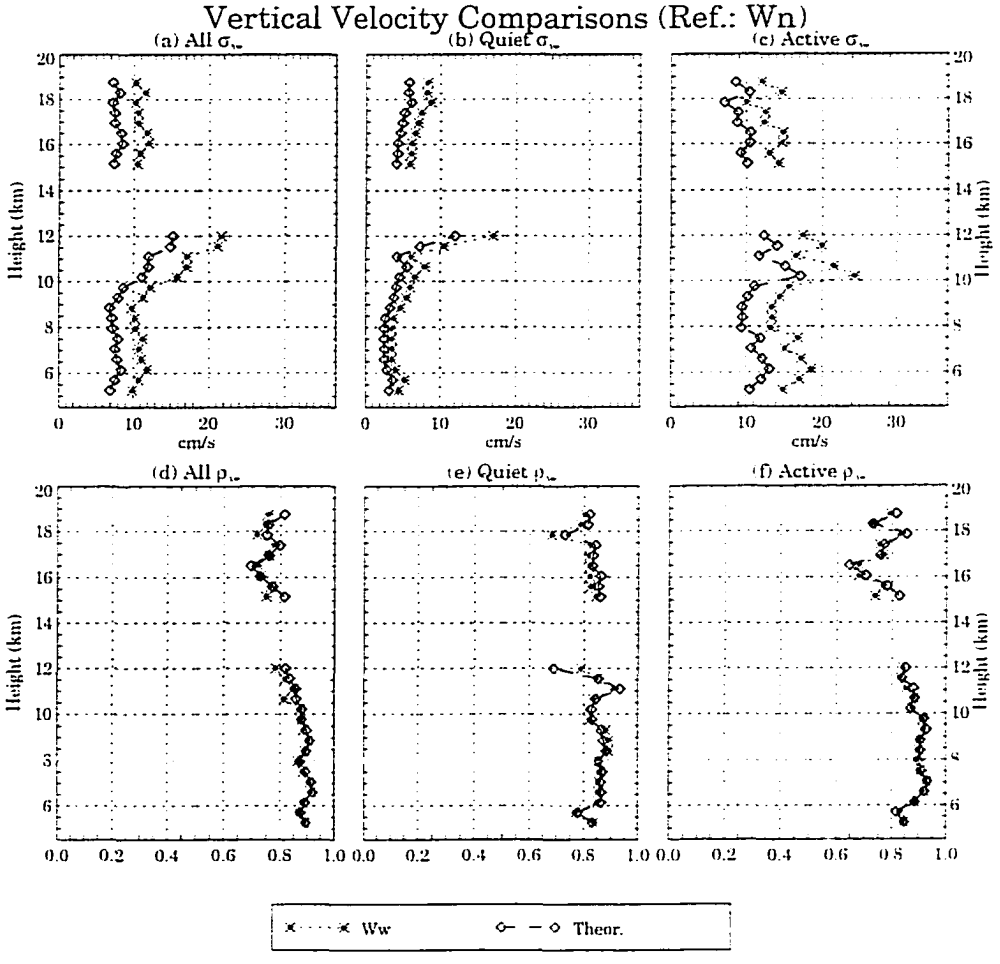


Figure 6.10. Standard deviations of the differences ($\sigma_{\Delta w}$ by stars in upper panels) and correlation coefficients ($\rho_{\Delta w}$ by stars in lower panels). In addition to the standard "errors" of vertical velocities (σ_ϵ by diamonds in first column) and theoretical-derived correlation coefficients ($\hat{\rho}_{\Delta w}$).

common mean vertical velocity, Δw is a constant difference, w is the "true" geophysical vertical velocity perturbation, and x and y are the "statistical" errors in the narrow and wide beam, respectively. The geophysical (w) and "statistical" (x and y) components have been assumed to be uncorrelated zero-mean Gaussian random processes with variances $\hat{\sigma}_w^2$, σ_x^2 and σ_y^2, respectively. Furthermore, the "statistical" variances of the two measurements are assumed to be equal, i.e., $\sigma_x^2 = \sigma_y^2 = \sigma_\epsilon^2$. The geophysical variance ($\hat{\sigma}_w^2$) is obtained from the difference between the narrow beam variance (σ_w^2 for Wn in Figure 6.6) and the "statistical" variance (σ_ϵ^2). A similar approach can be applied to the other vertical velocity techniques.

Given the good agreement between the measured and the theoretical correlation coefficients, approximated standard "errors" in the vertical velocities are between 2 cm s^{-1} and 5 cm s^{-1} during the quiet period, and $\approx$ 10 cm s^{-1} during the active period.

6.5 Discussion

This section begins by discussing the validity of using the narrow beam Zenith technique (Wn) as a reference for vertical velocity comparisons. Consider the following points:

• Theoretically the mean vertical velocity, $\bar{w}$, should go to zero at the tropopause in response to the reversal of the quasi horizontal temperature gradient. Moreover, even though the Jicamarca radar is located in a mountainous region, one would expect that under quiet conditions $\bar{w} \approx 0$. Finally, under active conditions, there could be a gravity wave with a vertical component of propagation. Such a wave, due to orographic effects, would decay in intensity with increasing height. This wave could have several levels of non-divergence (LNDs) without violating any physical principle (e.g. Bluestein [1992]). All of these statements appear to be best matched to the Wn profiles in Figure 6.9.

• The SNR of the narrow beam antenna (E) should be higher than the SNR obtained using the other antennas. It follows that Wn measurements will show less statistical uncertainties than the other w measurements (e.g. Woodman [1985]). The results shown in Figure 6.2 support this statement.

On the basis of these arguments we consider that Wn is a reasonably good reference profile for our comparisons. We now proceed to discuss the results obtained using the other techniques.

• Wide beam results (Ww). The profiles shown in Figure 6.8 provide a consistent picture of the mean relative differences between Ww and the reference profile. Specifically, the Figure 6.8 profiles show consistent negative values below 12 km and equally consistent positive values above 15 km. Presumably, most of these discrepancies are beam-width dependent. Moreover, in most of the heights these mean differences are proportional to the variability of the vertical wind. Above 15 km. there are also discrepancies apparent in the "instantaneous" values given by the relatively low correlation coefficients (Figure 6.8).

We have also examined the possibility that the wide and narrow beam discrepancies are due to the type of estimator used, and conclude that this is not the case. Different velocity estimators have been applied to both direct measurements using three ground clutter removal procedures, two types of spectral windows (Hanning and rectangular), and two moment estimators (NOAA's algorithm, Gaussian fitting). The greatest mean difference between estimators was less than 0.5 cm s^{-1}.

Similar patterns to the wide beam results, i.e., sinking in the troposphere and rising in the stratosphere, have been obtained at the tropics (Gage et al. [1991b]; Huaman and Balsley [1996]) and at mid-latitudes (Fukao et al. [1991]; Yoe et al. [1992]; Nastrom and VanZandt [1994]). The results at the tropics, however, are of smaller magnitude than the wide beam results, but the beam-widths in the tropics are also smaller ($\approx 2°$). On the other hand, the results at mid-latitudes are of similar magnitude to

the wide beam results at Jicamarca. Recently, Muschinski [1996] has theoretically associated the effects of Kelvin-Helmholtz Instability (KHI) to the mean vertical velocity biases seen at mid-latitudes. The "KHI bias" is downward in the shear zone below the horizontal wind speed maximum of a jet stream and upward above. Muschinski has also suggested that the direction of the AOAs is defined by the sign of $\partial u/\partial z$. In our case, the wide beam vertical velocity profile changes direction around 15 km (Figure 6.9), while the horizontal wind speed maximum is above 17 km (Figure 6.4). Our median values of AOAs (Figure 6.5), however, appear to be uncorrelated with the sign of $\partial u/\partial z$ (Figure 6.4). From these statements, KHI effects alone are not the cause of the Wn and Ww discrepancies.

• "Quasi" DBS results" (PSS). In contrast to the wide beam results, PSS/Wn comparisons show better agreement in the mean values. It appears that by sampling different pointing directions (nine), we have decreased the differences present in the Ww results, particularly during quiet conditions. The PSS "instantaneous" results, on the other hand, show strong similarities to the Ww results.

• "Classical" SA results (FCA-type). Analysis of the "true" velocity estimators (FCAM, FCAB, FCAH and FSAM) shows good agreement among their measurements in the lower troposphere. This agreement corresponds to: (a) similar correlation coefficients ($\rho_{\Delta w} > 0.8$), (b) similar biases (within ± 0.5 cm s^{-1}), and (c) similar standard "errors". Most of the discrepancies between the FCA-type estimators are seen above 15 km. These results are consistent with the discrepancies in horizontal velocity estimates discussed in Chapter 4.

Comparing the radial velocity results (FCAr) of Figure 6.8, with the $\hat{w}$ results of Figure 6.7, we see that a major improvement is achieved using the corrected values (higher correlations, lower standard "errors" and lower "biases") than by using the radial velocities. Below 12 km, the biases given by the FCA-type techniques are less than those given by Ww, but still ≈ -1 cm s^{-1} and ≈ -4 cm s^{-1} for the quiet

and active period, respectively. Moreover, at these heights, the FCA-type results are closer to the Ww results than to the Wn results. Above 15 km, $\tilde{w}$ results are very similar to Ww results, i.e. both exhibit similar rising motions.

Even though we have shown improvements of corrected vertical velocities relative to those obtained using direct measurements on the vertical wide beam, we have the following concerns with these corrections:

• Horizontal velocity percentage of failures. Since the corrections are done in a SA system, the horizontal velocity is obtained from a SA technique. However, the percentage of failures of the SA horizontal velocity techniques are > 10% (see Chapter 4). This failure rate causes a similar number of failures on the corrected vertical velocities. While these "gaps" are not so important in determining mean values of w, short-term spectral studies (e.g. gravity wave parameterization) will be strongly affected. This problem could be reduced by using other horizontal velocity estimates, e.g., the horizontal wind estimates from a simultaneous DBS configuration.

• Horizontal velocity magnitude. From the results presented in Chapter 4, and also in Section 6.4, we see that most SA "true" velocity estimators give close to the same horizontal values at lower tropospheric heights. Above 15 km, however, the horizontal velocity estimates using these different techniques vary considerably. In general, with respect to the horizontal velocity reference (FCAM), (a) FSAM estimates give underestimated values while (b) FCAB and FCAH estimates give overestimated values. As a result, the different techniques produce different corrected vertical velocities. Coincidentally, these heights are characterized by lower SNR and more aspect sensitivity than the tropospheric heights.

• Angle-of-arrival statistical errors. Typical statistical errors in AOA measurements are $\approx 0.15°$ for 3-minute estimates (see Chapter 5). The magnitude of these errors decreases with increasing integration time. In long-term measurements, these errors should average out. However, in studies using short-term w measurements

(< 15 min.) these AOA errors can be significant, particularly during quiet periods when the w oscillations are low ($\sigma_w < 8$ cm s^{-1}). For example, a 20 m s^{-1} horizontal wind velocity would produce a vertical velocity error of 5 cm s^{-1}, a value comparable to the w fluctuations.

Although, in general, the corrected vertical velocities (for off-vertical effective pointing) compare better than the Ww velocities, with the Wn reference , they still show significant mean differences ($\approx$ -1 to -4 cm s^{-1} below 12 km and $\approx$ 3 to 8 cm s^{-1} above 15 km). Assuming that our reference is reasonable, there must be factors other than off-vertical effective pointing that contribute to biases in the wide beam techniques. It is important to stress that SA techniques also use wide beam antennas. Other possible factors that could contribute to the vertical velocity mean differences could be due to:

• Gravity waves. Nastrom and VanZandt [1994] suggested that downward-moving portion of gravity waves with a vertical component of propagation would exhibit larger reflectivity than the upward moving portion. Ww and the corrected vertical velocities seem to be in agreement with this theory, i.e. they show stronger downward velocities than our reference. However, this theory does not explain why Wn and PSS results do not show such pronounced downward velocities.

• Wind-field gradients. Larsen and Palmer [1997] and Palmer et al. [1997a] showed how gradients in the wind field could affect vertical velocity and AOA estimates obtained from a finite beam-width radar. Divergence effects depend on the beam-width, so that the smaller the beam-width the smaller the errors on w due to divergence. This process would explain differences between Wn and Ww measurements. It would also explain the closer agreement between corrected vertical velocities ($\hat{w}$) and Ww, relative to the poorer agreement between $\hat{w}$ and Wn. In addition, if vorticity is playing a major role, AOAs will be contaminated. This would produce biases in $\hat{w}$ estimates (Larsen and Palmer [1997]). The PSS results are explained if the

wind-field gradient effects depend on zenith angle, since those effects appear to be averaged out in our nine-beam configuration. Assuming that approximate values of horizontal divergence can be obtained from $\partial w/\partial z$ (continuity equation), profiles of the wide beam mean differences should change sign during the active period, particularly below 12 km. However, our observations do not display such a change of sign. We are not able to quantify the effects of gradients in the wind-field because we did not have other means to measure divergence and vorticity at the time of the experiment.

Examining Figure 6.9, we can see that the 5-day mean average has influences from the active mean values. These active mean values are almost certainly caused by orographic effects. Therefore, if validity of Wn measurements has to be tested against vertical velocities obtained by operational analysis, one would need quiet conditions to avoid this "meteorological" noise (Nastrom et al. [1985]).

6.6 Conclusions

We have shown that the choice of the narrow beam vertical velocity profile is a reasonable reference (Wn) for comparing a variety of vertical velocity measurement techniques. Our choice was based on the reasonable value of the observed mean vertical velocity profile, and on the high SNR signals available at almost all heights using this technique. While this type of measurement appears to give the most reasonable vertical velocity results, implementation of the Wn technique is very expensive. It is important to stress that the most reasonable profile is not necessarily the correct profile. In lieu of a true "standard" for comparison, it will be ultimately necessary to compare radar results with other methods. Currently the best comparison appears to lie with the synoptic-scale vertical velocity obtained from observational analysis.

The wide beam results (Ww) exhibit substantial discrepancies, primarily

in the mean values relative to the Wn profiles. The Ww show downward mean differences at tropospheric heights, and upward mean differences above 15 km. On the other hand, the validity of "instantaneous" vertical velocities obtained by wide beams are somewhat guarantee. Both wide and narrow beam Zenith techniques show very similar values of variances. Given the high correlation coefficients between both techniques, the small discrepancies in the instantaneous values are almost certainly of statistical nature. The statistical "errors" are < 5 cm s^{-1} during quiet conditions and ≈ 10 cm s^{-1} during active conditions. Similar conclusions can be inferred for the "instantaneous" vertical velocities of the other techniques, particularly below 12 km.

The PSS results are also very encouraging. Even though the PSS technique is supposed to be affected by anisotropic scattering, PSS results show better agreement with the reference than either the vertical wide beam profiles or the "corrected" profiles.

The concept of correcting vertical velocities does appear to work. Using the SA configuration, we were able to obtain "corrected" vertical velocities by compensating our "off-vertical" echo returns by subtracting out the horizontal velocity contributions. These off-vertical returns include a systematic contribution ("off-vertical" angle $> 0.46°$) due to the antenna pointing directions and the geometry of the SA system, as well as atmospheric contributions that can have instantaneous off-vertical values as high as 1°. While these corrected measurements give better mean value results than the wide beam Zenith technique, the corrected mean values lie between the Wn and Ww results, and are close to Ww profile.

Based on the above discussions, vertical velocity measurements, particularly mean values, appear to be affected not only by "effective" beam pointing problems, but also by other effects that depend on the radar beam-width. Future comparisons of narrow beam and wide beam results will help to qualify and possible to quantify

these beam-width-depend effects.

CHAPTER 7

SUMMARY AND FUTURE WORK

In this thesis, we have examined a number of techniques for wind velocity estimation, and have put special emphasis on the vertical component of the wind. In the process of these examinations, we have studied the angle-of-arrival (AOA) measurements that are needed for a specific type of vertical velocity techniques.

In Chapter 4, we have shown that the simpler horizontal velocity techniques that assume horizontally isotropic scattering compare very closely with more complicated full correlation analysis techniques, particularly below 12 km. The assumption of a Gaussian form of the cross-correlation functions is demonstrated to be reasonable at high SNR tropospheric heights. In general, for all of the horizontal velocity techniques examined, there is excellent agreement in the wind direction. On the other hand, considerable discrepancies between techniques occur in the wind speed estimates, particularly above 15 km.

In Chapter 5, based on scattering theory, we have shown that the interpretation of AOA measurements has been improved. We have isolated the contributions from (1) tilted layers, (2) antenna pointing, and (3) geometrical contributions. It is also shown that the estimation of most AOA contributions, requires the measurement of the scattering parameters. In addition, we have shown that the correction of vertical velocities can be accomplished without explicitly determining the off-vertical atmospheric contributions. The comparisons of a time- and a frequency-domain method for AOA measurements showed that the time-domain method is easier to implement, faster to calculate, and more reliable than the frequency-domain method. Assuming that the scattering, on average, is isotropic, AOA radar measurements are

in good agreement with theoretical values.

May [1993] suggested that collocated SA systems (where the transmitting antenna is centered on the array of receiving antennas) should be used in vertical velocity corrections instead of non-collocated configurations. However, using scattering theory and based on the same arguments used by May, we have shown in Section 5.3 that such corrections can be done even if one uses non-collocated systems. It is important, however, to correctly use the AOA and radial velocity information. Having said this, it still may be possible that there are other factors, which were not considered in our development, that make non-collocated systems unsuitable for vertical velocity corrections.

The choice of the narrow beam vertical velocity (Wn) profile as a reference (Chapter 6) has been shown to be reasonable for comparing concurrent vertical velocity measurements obtained by a variety of techniques. Our choice to use the narrow beam as a reference was based on the reasonable value of the mean of the observed vertical velocity profile, and the high SNR signals available at almost all heights using this technique. While the narrow beam measurement appears to give the most reasonable vertical velocity results, implementation of the technique is very expensive and more feasible alternatives should be considered.

Comparison of the vertical velocity techniques under different scenarios, shown in Table 7.1, is summarized with a subjective rating. We rate the following vertical velocity techniques: narrow beam Zenith technique (Wn), wide beam Zenith technique (Ww), "quasi" DBS technique (PSS) and the "classical" SA technique ($\bar{w}$). Our ratings are based on the results presented in Chapter 6. Specifically, the ratings on mean vertical velocities come from the results presented in Figure 6.9. The key for these ratings is: poor ($\star$), moderate ($\star\star$), good ($\star\star\star$), excellent ($\star\star\star\star$) and perfect ($\star\star\star\star\star$). Examination of Table 7.1 shows that the overall rating for each technique is: excellent for the narrow beam Zenith technique, moderate for the wide

beam Zenith technique, good for the "quasi" DBS technique, and moderate for the "classical" SA techniques.

Table 7.1. Subjective ratings for vertical velocity techniques. [Wn: narrow beam Zenith, Ww: wide beam Zenith, PSS: "quasi" DBS, and $\hat{w}$: "classical" SA]

	Wn	Ww	PSS	$\hat{w}$
Below 12 km	★★★★	★★	★★★	★★★
Above 15 km	★★★★	★★	★★★	★
Quiet conditions	★★★★	★★	★★★	★★
Active conditions	★★★	★★	★★	★★
Mean values	★★★★	★	★★★	★★
3-minute values	★★★	★★★	★★	★
On-line processing	★★★★	★★★★	★★★	★★
Overall	★★★★	★★	★★★	★★

It is important to stress here that an excellent profile rating does not necessarily imply that the profile is correct. In lieu of a true "standard" for comparison, it ultimately will be necessary to compare radar results with other methods. Currently the best comparison appears to lie with the synoptic-scale vertical velocity obtained from observational analysis.

Improvements to three-dimensional wind velocity measurement comparisons could be achieved in the future with the following activities:

• More narrow-wide beam comparisons. The cause of the discrepancies between narrow and wide beam vertical velocities is still not fully understood. Complementary measurements of divergence and vorticity by observational analysis could help to quantify their effects on vertical velocity measurements. A climatology of the narrow beam measurements could help to validate better these type of measurements. Nowadays, these type of measurements are only possible using the Jicamarca Radar.

• Concurrent collocated and non-collocated SA measurements. These types of measurements could help to confirm the assumptions used in Chapter 5. These additional measurements would give more inputs on the understanding of the actual scattering process. This configuration could be achieved using either the Jicamarca Radar or

the MU radar.

• Experiments using higher sensitivity. Among the various techniques studied throughout this thesis, most of the discrepancies occur above 12 km, where SNR values are low. The use of more power (> 100 kW) or pulse-coding would help to increase the SNR at these heights. This improvement would, in turn, increase the understanding of the observed discrepancies, particularly in the case of the SA techniques. Pulse-coding was not implemented in the data used in this work. This choice enabled us to have the lowest possible height coverage. These recommended experiments can be done by any SA VHF system with high power capabilities.

• Concurrent Full-DBS and SA measurements. The horizontal winds obtained with the Full-DBS configuration can be used to get corrected vertical velocities. This procedure decreases the percentage of gaps on such vertical velocities. As mentioned in Chapter 3, the Jicamarca radar is not suitable for Full-DBS measurements. This is because the maximum oblique angles are small, so horizontal velocity estimates are affected by aspect sensitivity (Tsuda et al. [1986]). The MU radar is an appropriate choice for these type of measurements.

• Scattering theory and the use of multiple receiving antennas. With some modifications to the scattering model used in Chapter 5, one can study the feasibility of using at least four non-coplanar receiving antennas for vertical velocity measurements. Moreover, the use of multiple antennas (coplanar or non-coplanar) allows the relaxation of some simplifying assumptions. For example, one can then assume that the spatial spectrum of the refractive index field is composed of more than one Gaussian "blob" (Woodman 1997, personal communication). Not only the wind-velocity estimation could be improved with this approach but also the "imaging" of the atmospheric scattering mechanisms could be achieved. We have already collected data with a four-antenna and a seven-antenna configuration. The processing of these data will be done in the near future. In addition, the PSS pointing can be

more precisely determined not only by incorporating aspect sensitivity parameters but also by using more than three receiving antennas.

A further benefit of using scattering theory is that one can study the feasibility of using rectangular antennas in SA techniques. For example, it is possible to make an SA modification to existing wind profilers that have two collocated COCO (coaxial collinear) antennas. This modification could split each antenna in halves, so that the resultant SA system would be composed of four rectangular antennas. The same array could be used simultaneously in the normal DBS mode, without loss of data.

• Evaluation of wind velocity estimators using simulated data. The precision of each of the wind velocity techniques has been a concern throughout this work. A careful evaluation of the different techniques using simulated data could better quantify the statistical uncertainties expected for each technique (May [1988]; Frehlich et al. [1994]). Not only can the statistical errors on the velocity estimators be obtained but also the statistical errors in AOAs and scattering parameters. Moreover, the generation of simulated data can be greatly improved by incorporating theoretical cross-correlation functions obtained with the scattering theory mentioned above.

• Scattering modeling. A simple computer model similar to the one presented by Holdsworth and Reid [1995] and Chau et al. [1995], could help to implement components that are not easy to incorporate on the scattering theory or simulated data. For example, one can incorporate a monochromatic specular wave in the model to study its effects on the different velocity estimators.

BIBLIOGRAPHY

Armstrong, J. W., and W. A. Coles, Analysis of three-station interplanetary scintillation, **J. Geophys. Res.**, 77(25), 4602–4610, 1972.

Balsley, B., and R. Garello, The kinetic energy density in the troposphere, stratosphere, and mesosphere: a preliminary study using the Poker Flat MST radar in Alaska, **Radio Sci.**, 20, 1355–1361, 1985.

Balsley, B. B., The MST technique - A brief review, **J. Atmos. Terr. Phys.**, 43, 495–509, 1981.

Balsley, B. B., and D. A. Carter, Mountain waves in the tropical Pacific atmosphere: A comparison of vertical wind fluctuations over Pohnpei and Christmas Island using VHF wind profilers, **J. Atm. Sci.**, 46, 2698–2715, 1989.

Balsley, B. B., and A. C. Riddle, Monthly mean values of the mesospheric wind field over Poker Flat, Alaska, **J. Atm. Sci.**, 41, 2368–2375, 1984.

Balsley, B. B., W. L. Ecklund, D. A. Carter, A. C. Riddle, and K. S. Gage, Average vertical motions in the tropical atmosphere observed by a radar wind profiler on Pohnpei (7°N latitude, 157°E longitude), **J. Atm. Sci.**, 45(3), 396–405, 1988.

Barnes, S. L., On the accuracy of Omega diagnostic computations, **Mon. Weather Rev.**, 114, 1664–1680, 1986.

Battan, L. J., **Fundamentals of Meteorology**, Prentice-Hall, Inc., 1979.

Belt, C. L., and H. E. Fuelberg, The effects of random errors in rawinsonde data on derived kinematic quantities, **Mon. Weather Rev.**, 110, 91–101, 1982.

Bluestein, H. B., **Synoptic-Dynamic meteorology in midlatitudes**, Oxford University Press, 1992.

Briggs, B. H., Radar observations of atmospheric winds and turbulence: a comparison of techniques, **J. Atmos. Terr. Phys.**, 42, 823–833, 1980.

Briggs, B. H., The analysis of spaced sensors records by correlation techniques, in **Handbook for the Middle Atmosphere Program**, pp. 166–186, University of Illinois, Urbana, Special Committee for Solar-Terrestrial Physics Secretariat, 1984.

Briggs, B. H., and A. D. Maude, Spaced sensor observations of pattern motion, **J. Geophys. Res.**, **83**(A11), 5309–5311, 1978.

Briggs, B. H., and R. A. Vincent, Spaced-antenna analysis in the frequency domain, **Radio Sci.**, **27**(2), 117–129, 1992.

Briggs, B. H., G. J. Phillips, and D. H. Shinn, The analysis of observations on spaced receivers of the fading radio signals., **Proc. Phys. Soc. London. 63**, 106–121, 1950.

Brown, W. O., G. J. Fraser, S. Fukao, and M. Yamamoto, Spaced antenna and interferometric velocity measurements with MF and VHF radars, **Radio Sci.**, **30**(4), 1281–1292, 1995.

Chau, J. L., B. B. Balsley, and L. E. Liziola, Preliminary results using a simple scattering model, Extended abstract and poster, Ninth International Symposium on equatorial aeronomy, Bali, Indonesia, 1995.

Cohn, S. A., C. L. Holloway, S. Oncley, R. Doviak, and R. Lataitis, Validation of a UHF spaced antenna wind profiler for high resolution boundary layer observations, **Radio Sci.**, **32**(3), 1279–1296, 1997.

Currier, P. E., S. A. Avery, B. B. Balsley, K. S. Gage, and W. L. Ecklund, Combined use of 50 MHz wind profilers in the estimation of raindrop size distributions, **Geophys. Res. Lett.**, **19**, 1017–1020, 1992.

Doviak, R., and D. S. Zrnić, **Doppler radar and weather observations**, Academic Press, Inc., second edn., 1993.

Doviak, R. J., R. J. Lataitis, and C. L. Holloway, Cross correlations and cross spectra for spaced antenna wind profilers, **Radio Sci.**, **31**(1), 157–180, 1996.

Dunn, L. B., Evaluation of vertical motions: Past, present, and future, **Wea. Forecasting**, **6**, 65–75, 1991.

Durran, D. R., and L. W. Snellman, The diagnosis of synoptic-scale vertical motion in an operational environment, **Wea. Forecasting**, **2**, 17–31, 1987.

Ecklund, W. L., B. B. Balsley, D. A. Carter, A. C. Riddle, M. Crochet, and R. Garello, Observations of vertical motios in the troposphere and lower stratosphere using three closely spaced ST radars, **Radio Sci.**, **20**(6), 1196–1206, 1985.

Franke, P. M., D. Thorsen, M. Champion, S. J. Franke, and E. Kudeki, Comparison of time- and frequency-domain techniques for wind velocity estimation using multiple-receiver MF radar data, **Geophys. Res. Lett.**, **17**, 2193–2196, 1990.

Frehlich, R., S. M. Hannon, and S. W. Henderson, Performance of a 2-μm coherent Doppler lidar for wind measurements, **J. Atmos. Oceanic Technol.**, 11(6), 1517–1528, 1994.

Fritts, D. C., and Z. Luo, Dynamical and radiative forcing of the summer mesospause circulation and thermal structure I. mean solstice conditions, **J. Geophys. Res.**, **100**, 3119–3128, 1995.

Fukao, S., T. Sato, T. Tsuda, M. Yamamoto, M. D. Yamanaka, and S. Kato, MU radar: New capabilities and system calibrations, **Radio Sci.**, **25**(4), 477–485, 1990.

Fukao, S., M. F. Larsen, M. D. Yamanaka, H. Furukawa, T. Tsuda, and S. Kato, Observations of a reversal in long-term average vertical velocities near the jet stream wind maximum, **Mon. Weather Rev.**, **119**, 1479–1489, 1991.

Gage, K. S., Radar observations of the free atmosphere structure and dynamics, in **Radar in Meteorology**, edited by D. Atlas, pp. 534–565, Amer. Meteor. Soc., 1990.

Gage, K. S., and G. D. Nastrom, A simple model for the enhanced frequency spectrum of vertical velocity based on tilting of atmospheric layers by lee waves, **Radio Sci.**, **25**(5), 1049–1056, 1990.

Gage, K. S., D. A. Carter, and W. L. Ecklund, The effect of gravity waves on specular echoes observed by the Poker Flat MST radar, **Geophys. Res. Lett.**, **8**(6), 599–602, 1981.

Gage, K. S., J. R. MzAfee, W. G. Collins, D. Söderman, H. Böttger, A. Radford, and B. B. Balsley, A comparison of winds observed at Christmas Island using a wind profiler Doppler radar with NMC and ECMWF analyses, **Bull. Am. Meteorol. Soc.**, **69**, 1041–1057, 1988.

Gage, K. S., J. R. McAfee, D. A. Carter, W. L. Ecklund, and A. C. R. an G. C., Long-term mean vertical motion over the tropical Pacific: Wind-profiling Doppler radar measurements, **Science**, **254**, 1771–1773, 1991a.

Gage, K. S., B. B. Balsley, W. L. Ecklund, D. A. Carter, and J. R. McAfee, Wind profiler-related research in the tropical Pacific, **J. Geophys. Res.**, **96**(Supplement), 3209–3220, 1991b.

Gage, K. S., J. R. McAfee, and G. C. Reid, Diurnal variation in vertical motion over the central equatorial Pacific from VHF wind-profiling Doppler radar observations at Christmas Island (2°N, 157°W), **Geophys. Res. Lett.**, **19**, 1827–1830, 1992.

Green, J. L., and K. S. Gage, Observations of stable layers in the troposphere and stratosphere using VHF radar, **Radio Sci.**, **15**(2), 395–405, 1980.

Hardy, K. R., and K. S. Gage, The history of radar studies of the clear atmosphere, in **Radar in Meteorology**, edited by D. Atlas, pp. 534–565, Amer. Meteor. Soc., 1990.

Hildebrand, P. H., and R. S. Sekhon, Objective determination of the noise level in Doppler spectra, **J. Appl. Meteor.**, **13**, 808–811, 1974.

Hocking, W. K., P. May, and J. Röttger, Interpretation, reliability and accuracies of parameters deduced by spaced antenna method in middle atmosphere applications, **Pure Appl. Geophysics**, **130**(2/3), 571–604, 1989.

Holdsworth, D. A., An investigation of biases in the full spectral analysis technique, **Radio Sci.**, **32**(2), 769–782, 1997.

Holdsworth, D. A., and I. M. Reid, A simple model of atmospheric radar backscatter: Description and application to the full correlation analysis of spaced antenna data, **Radio Sci.**, **30**(4), 1263–1280, 1995.

Holloway, C., R. Doviak, S. Cohn, R. Lataitis, and J. V. Baelen, Cross correlations and cross spectra for spaced antenna wind profilers 2. Algorithms to estimate wind and turbulence, **Radio Sci.**, **32**(3), 967–982, 1997a.

Holloway, C., R. Doviak, and S. Cohn, Cross correlations of fields scattered by horizontally anisotropic refractive index irregularities, **Radio Sci.**, **32**(5), 1911–1920, 1997b.

Holton, J. R., **An introduction to dynamic meteorology**, Academic Press, Inc., third edn., 1992.

Hoppe, U. P., and D. C. Fritts, High-resolution measurements of vertical velocity with the european incoherent scatter VHF radar 1. Motion field characteristics and measurements biases, **J. Geophys. Res.**, **100**(D8), 16813–16825, 1995.

Houze, R. A., Observed structure of mesoscale convective systems and implications for large-scale heating, **Q. J. R. Meteorol. Soc.**, **115**, 425–461, 1989.

Huaman, M. M., and B. Balsley, Long-term average vertical motions observed by VHF wind profilers: The effect of slight antenna-pointing inaccuracies, **J. Atmos. Oceanic Technol.**, **13**(3), 560–569, 1996.

Kudeki, E., Spaced antenna and interferometry methods, Second International School of Atmospheric radar, Lecture notes, 1995.

Kudeki, E., and R. F. Woodman, A postatistics steering technique for MST radar applications, **Radio Sci.**, **25**(4), 591–594, 1990.

Larsen, M. F., and R. D. Palmer, A relationship between horizontal flow gradients, in-beam incidence angles, and vertical velocities, **Radio Sci.**, **32**(3), 1269–1277, 1997.

Larsen, M. F., and J. Röttger, The spaced antenna technique for radar wind profiling, **J. Atmos. Oceanic Technol.**, **6**, 920–938, 1989.

Larsen, M. F., and J. Röttger, VHF radar measurements of in-beam incidence angles and associated vertical-beam radial velocity corrections, **J. Atmos. Oceanic Technol.**, **8**, 477–490, 1991.

Larsen, M. F., J. Röttger, and T. S. Dennis, A comparison of operational analysis and VHF wind profiler vertical velocities, **Mon. Weather Rev.**, **116**, 48–59, 1988.

Larsen, M. F., S. Fukao, M. D. Yamanaka, T. Tsuda, and S. Kato, A comparison of VHF radar vertical-velocity measurements by a direct vertical-beam method and by a VAD technique, **J. Atmos. Oceanic Technol.**, **8**, 766–776, 1991.

Larsen, M. F., R. D. Palmer, S. Fukao, R. F. Woodman, M. Yamamoto, T. Tsuda, and S. Kato, An analysis technique for deriving vector winds and in-beam incidence angles from radar interferometer measurements, **J. Atmos. Oceanic Technol.**, **9**(1), 3–14, 1992.

Lee, J. L., G. L. Browning, and Y. F. Xie, Estimating divergence and vorticity from the wind profiler network hourly wind measurements, **Tellus, Ser. A. Dyn. meteorol. oceanogr.**, **47A**, 892–910, 1995.

Liu, C. H., and C. J. Pan, New observational techniques for studying the dynamics of the middle atmosphere using the Chung Li VHF radar, **J. Atmos. Terr. Phys.**, **55**(7), 1055–1066, 1993.

Liziola, L. E., and B. B. Balsley, Horizontally propagating quasi-sinusoidal tropospheric waves observed in the lee of the Andes, **Geophys. Res. Lett.**, **24**(9), 1075–1078, 1997.

Low, D. J., Studies of the lower atmosphere with a VHF wind profiler, Ph.D. thesis, University of Adelaide, Australia, 1996.

Mace, G. G., and T. P. Ackerman, Assesment of error in synoptic-scale diagnostics derived from wind profiler and radiosonde network data, **Mon. Weather Rev.**, **124**, 1521–1534, 1996.

Mace, G. G., D. O. Starr, T. P. Ackerman, and P. Minnis, Examination of coupling between an upper-tropospheric cloud system and synoptic-scale diagnosed from wind profiler and radiosonde data, **J. Atm. Sci.**, **52**(23), 4094–4127, 1995.

Mattocks, C., and R. Bleck, Jet streak dynamics and geostrophic adjustments processes during the initial stages of lee cyclogenesis, **Mon. Weather Rev.**, **114**, 2033–2056, 1986.

May, P., and R. G. Strauch, Reducing the effect of ground clutter on wind profiler velocity measurements, 1997.

May, P. T., Statistical errors in the determination of wind velocities by the spaced antenna technique, **J. Atmos. Terr. Phys.**, **50**(1), 21–32, 1988.

May, P. T., Comments on "VHF radar measurements of in-beam incidence angles and associated vertical-beam radial velocity corrections, **J. Atmos. Oceanic Technol.**, **10**, 432–434, 1993.

May, P. T., and D. K. Rajopadhyaya, Wind profiler observations of vertical motion and precipitation microphysics of a tropical squall line, **Mon. Weather Rev.**, **124**(4), 621–633, 1996.

May, P. T., and R. G. Strauch, An examination of wind profiler signal processing algorithms, **J. Atmos. Oceanic Technol.**, **6**, 731–735, 1989.

May, P. T., T. Sato, M. Yamamoto, S. Kato, T. Tsuda, and S. Fukao, Errors in the determination of wind speed by Doppler radar, **J. Atmos. Oceanic Technol.**, **6**, 235–242, 1989.

McAfee, J. R., K. S. Gage, and R. G. Strauch, Vertical velocities at Platteville Colorado: An intercomparison of simultaneous measurements by VHF and UHF profilers, **Radio Sci.**, **34**(4), 1027–1042, 1995.

Meek, C. E., An efficient method for analysing ionospheric drifts data, **J. Atmos. Terr. Phys.**, **42**, 835–839, 1980.

Meek, C. E., and A. H. Manson, Angle-off-arrival oscillations in the mesosphere as seen by medium frequency (MF) radar, **J. Atmos. Terr. Phys.**, **54**(3/4), 277–293, 1992.

Muschinski, A., Possible effect of Kelvin-Helmholtz instability on VHF radar observations of the mean vertical wind, **J. Appl. Meteor.**, **35**, 2210–2217, 1996.

Nastrom, G. D., Detection of synoptic-scale vertical velocities using a MST radar, **Geophys. Res. Lett.**, **11**(1), 57–60, 1984.

Nastrom, G. D., and T. E. VanZandt. Mean vertical motions seen by radar wind profilers. **J. Appl. Meteor.**, **33**(8), 984-995, 1994.

Nastrom, G. D., and T. E. VanZandt, Biases due to gravity waves in wind profiler measurements of winds. **J. Appl. Meteor.**, **35**, 243–257, 1996.

Nastrom, G. D., and J. M. Warnock, Vertical motions estimated using data from a single station and a form of the adiabatic method, **J. Appl. Meteor.**, **33**, 65–73, 1994.

Nastrom, G. D., W. L. Ecklund, and K. S. Gage. Direct measurements of large-scale vertical velocities using clear-air Doppler radars. **Mon. Weather Rev.**, **113**, 708–718, 1985.

Nastrom, G. D., M. R. Peterson, J. L. G. K. S. Gage, and T. E. VanZandt. Sources of gravity wave activity seen in the vertical velocties observed by the Flatland VHF radar, **J. Appl. Meteor.**, **28**, 783–792, 1990a.

Nastrom, G. D., K. S. Gage, and W. L. Ecklund, Uncertainties in estimates of the mean vertical velocity from MST radar observations. **Radio Sci.**, **25**(5), 933–940, 1990b.

Nastrom, G. D., W. L. Clark, K. S. Gage, T. E. VanZandt, J. M. Warnock, R. Creasey, and P. M. Pauley, Case studies of the vertical velocity seen by the Flatland radar compared with indirectly computed values. **J. Atmos. Oceanic Technol.**, **11**, 14–21, 1994.

O'Brien, J. J., Alternative solutions to the classical vertical velocity problem. **J. Appl. Meteor.**, **9**, 197–203, 1970.

Palmer, R. D., R. F. Woodman, S. Fukao, T. Tsuda, and S. Kato, Three-antenna postatistic steering using the MU radar, **Radio Sci.**, **25**(6), 1105–1110, 1990.

Palmer, R. D., M. F. Larsen, R. F. Woodman, S. Fukao, M. Yamamoto, T. Tsuda, and S. Kato, VHF radar interferometry measurements of vertical velocity and the effect of tilted refractivity surfaces on standard Doppler measurements. **Radio Sci.**, **26**(2), 417–427, 1991.

Palmer, R. D., M. F. Larsen, E. L. Sheppard, S. Fukao, M. Yamamoto, T. Tsuda, and S. Kato, Postatistic steering wind estimation in the troposphere and lower stratosphere, **Radio Sci.**, **28**(3), 261–271, 1993.

Palmer, R. D., S. Vangal, M. F. Larsen, S. Fukao, T. Nakamura, and M. Yamamoto, Phase calibration of VHF spatial interferometry radars using stellar sources, **Radio Sci.**, **31**(1), 147–156, 1996.

Palmer, R. D., M. F. Larsen, and S. Vangal, Effects of finite beam width and wind field divergence on Doppler radar measurements: Simulations, **Radio Sci.**, **32**(3), 1179–1191, 1997a.

Palmer, R. D., M. F. Larsen, S. Fukao, and M. Yamamoto, On the relationship between aspect sensitivity and spatial interferometric in-beam incidence angles, **J. Atmos. Sol-. Terr. Phys.**, **in press**, 1997b.

Panofsky, B. A., Large-scale vertical velocity and divergence, **Compendium of Meteorology**, pp. 639-646, 1951.

Panofsky, H. A., Methods of computing vertical motion in the atmosphere, J. **Meteor.**, **3**, 45–49, 1946.

Pedder, M. A., On the errors of kinematic vertical motion estimation using divergence bias adjustment procedures, **Mon. Weather Rev.**, **109**, 1813–1816, 1981.

Peterson, V. L., and B. B. Balsley, Clear air Doppler radar measurements of the vertical component of wind velocity in the troposphere and stratosphere, **Geophys. Res. Lett.**, **6**(6), 933–936, 1979.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, **Numerical Recipes in C, The art of scientific computing**, Press Syndicate of the Univeristy of Cambridge, 40 West 20th Street, New York, NY 10011-4211, USA, second edn., 1992.

Ralph, F. M., M. Crochet, and S. V. Venkateswaran, Observations of a mesoscale ducted gravity wave, **J. Atm. Sci.**, **50**, 3277–3291, 1993.

Riddle, A. C., K. S. Gage, and B. B. Balsley, Practical use of a tropopause height determination algorithm on an MST radar data set, in **Handbook for the Middle Atmosphere Program**, **14**, pp. 64-73, University of Illinois, Urbana, Special Committee for Solar-Terrestrial Physics Secretariat, 1984.

Røyrvik, O., Spaced antenna drift at Jicamarca, mesospheric measurements., **Radio Sci.**, **18**(3), 461–476, 1983.

Röttger, J., Improvement of vertical velocity measurements, in **Handbook for the Middle Atmosphere Program**, **14**, pp. 164–173, University of Illinois, Urbana, Special Committee for Solar-Terrestrial Physics Secretariat, 1984.

Röttger, J., and H. M. Ierkic, Postset beam steering and interferometer applications to VHF radars to study winds, waves, and turbulence in the lower and middle atmosphere, **Radio Sci.**, **20**(6), 1461–1480, 1985.

Röttger, J., and M. F. Larsen, UHF/VHF radar techniques for atmospheric research and wind profiler applications, in **Radar in Meteorology**, edited by D. Atlas, pp. 253–281, Amer. Meteor. Soc., 1990.

Röttger, J., C. H. Liu, J. K. Chao, A. J. Chen, C. J. Pan, and I.-J. Fu, Spatial interferometer measurements with the Chung-Li VHF radar, **Radio Sci.**, **25**(4), 503–515, 1990a.

Röttger, J., C. L. Hoz, S. J. Franke, and C. H. Liu, Steepening of reflectivity structures detected in high-resolution Doppler spectra of polar mesosphere summer echoes (PMSE) observed with the ESICAT 224 MHz radar, **J. Atmos. Terr. Phys.**, **52**(10/11), 939–954, 1990b.

Sato, T., and R. F. Woodman, Spectral paramater estimation of CAT radar echoes in the presence of fading clutter., **Radio Sci.**, **17**, 817–826, 1982.

Sheppard, E. L., and M. F. Larsen, Analysis of model simulations of spaced antenna/radar interferometer measurements, **Radio Sci.**, **27**(5), 759–768, 1992.

Sheppard, E. L., M. F. Larsen, R. D. Palmer, S. Fukao, M. Yamamoto, T. Tsuda, and S. Kato, A statistical comparison of spaced and spatial interferometry wind estimation, **Radio Sci.**, **28**(4), 585–593, 1993.

Starr, D. O., and S. K. Cox, Cirrus clouds, part I: A cirrus cloud model, **J. Atm. Sci.**, **42**, 2663–2681, 1985.

Strauch, R. G., D. A. Merrit, P. Moran, K. B. Earnshaw, and D. van de Kamp, The Colorado wind-profiling network, **J. Atmos. Oceanic Technol.**, **1**, 37–49, 1984.

Sürücü, F., P. Franke, and E. Kudeki, On the use of software beam synthesis in multiple-receiver MF radar wind estimation, **Radio Sci.**, **27**(6), 775–782, 1992.

Thorsen, D., A statistical climatology of mesospheric gravity wave activity over Urbana, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1996.

Thorsen, D., S. J. Franke, and E. Kudeki, A new approach to MF radar interferometry for estimating mean winds and momentum flux, **Radio Sci.**, **32**(2), 707–726, 1997.

Tsuda, T., T. Sato, K. Hirose, S. Fukao, and S. Kato, MU radar observations of the aspect sensitivity of backscattered VHF echo power in the troposphere and lower stratosphere, **Radio Sci.**, **21**(6), 971–980, 1986.

Tsuda, T., W. E. Gordon, and H. Saito, Azimuth angle variations of specular reflection echoes in the lower atmosphere observed with the MU radar, **J. Atmos. Terr. Phys.**, **59**(7), 777–784, 1997a.

Tsuda, T., T. E. VanZandt, and H. Saito, Zenith-angle dependence of VHF specular reflection echoes in the lower atmosphere. **J. Atmos. Terr. Phys.**, **59**(7), 761–775, 1997b.

Van Baelen, J. S., and A. D. Richmond, Radar interferometry technique: Three-dimensional wind measurement theory, **Radio Sci.**, **26**(5), 1209–1218, 1991.

Van Baelen, J. S., A. D. Richmond, T. Tsuda, S. K. Avery, S. Kato, S. Fukao, and M. Yamamoto, Radar interferometry technique and anisotropy of the echo power distribution: First results. **Radio Sci.**, **26**(5), 1315–1326, 1991.

Vandepeer, B. G. W., and I. M. Reid, Some preliminary results obtained with a new Adelaide MF Doppler radar. **Radio Sci.**, **30**(4), 1191–1203, 1995.

VanZandt, T. E., G. D. Nastrom, and J. L. Green, Frequency spectra of vertical velocity from Flatland VHF radar data, **J. Geophys. Res.**, **96**(D2), 2845–2855, 1991.

Vincent, R. A., and I. M. Reid, HF Doppler measurements of mesospheric gravity wave momentum fluxes, **J. Atm. Sci.**, **40**(5), 1321–1333, 1983.

Vincent, R. A., P. T. May, W. K. Hocking, W. G. Elford, B. H. Candy, and B. H. Briggs, First results with the Adelaide VHF radar: spaced antenna studies of tropospheric winds, **J. Atmos. Terr. Phys.**, **49**(4), 353–366, 1987.

Weber, B. L., D. B. Wuertz, D. C. Law, A. S. Frisch, and J. M. Brown, Effects of small-scale vertical motion on radar measurements of wind and temperature profiles, **J. Atmos. Oceanic Technol.**, **9**(3), 193–209, 1992.

Woodman, R. F., Spectral moment estimation in MST radars. **Radio Sci.**, **20**(6), 1185–1195, 1985.

Woodman, R. F., On the true velocity in full correlation analysis. **Radio Sci.**, **30**(5), 1459–1465, 1995.

Woodman, R. F., and A. Guillén, Radar observations of winds and turbulence in the stratosphere and mesosphere, **J. Atm. Sci.**, **31**(2), 493–505, 1974.

Yamamoto, Y., T. Tsuda, and T. Adachi, Frequency spectra of wind velocity and temperature fluctuations observed with the MU radar-RASS. **Geophys. Res. Lett.**, **23**(4), 3647–3650, 1996.

Yoe, J. G., M. F. Larsen, and E. J. Zipser, VHF wind-profiler data quality and comparison of methods for deducing horizontal and vertical air motions in a mesoscale convective storm, **J. Atmos. Oceanic Technol.**, **9**(6), 713–727, 1992.

APPENDIX A

GROUND-CLUTTER TREATMENT

In this section, we present some algorithms that deal with constant or "slowly fading" ground-clutter signals. By "slowly fading" we mean a signal with a correlation time much larger than the correlation time of the atmospheric signal. Usually, this is the case for signals obtained by VHF radars. However, at UHF the fading of ground-clutter signals are often faster than at VHF, although they are still "slower" than the atmospheric signals. Methods dealing with these "faster fading" signals can be found in Sato and Woodman [1982] and May and Strauch [1997]. It is worth mentioning, that part of the "ground-clutter" signal could also come from the receivers and/or the digitization process, although these contributions are usually constant in time.

In Figure A.1 we show an example of a time-series of one of the complex receiving channels (R), for a single height. For the procedures we cover below, we have subdivided the time-series in n blocks (n = number of incoherent integrations). Each block having the same number of points (N_p) and separated by the vertical dotted lines. Also, we have calculated the mean value of each block ($\bar{R}_i$), denoted with solid circles. The same procedure is performed on the imaginary channel.

We have treated the ground-clutter signals in the following three ways: (1) time-domain subtraction, (2) frequency-domain subtraction, and (3) frequency-domain interpolation.

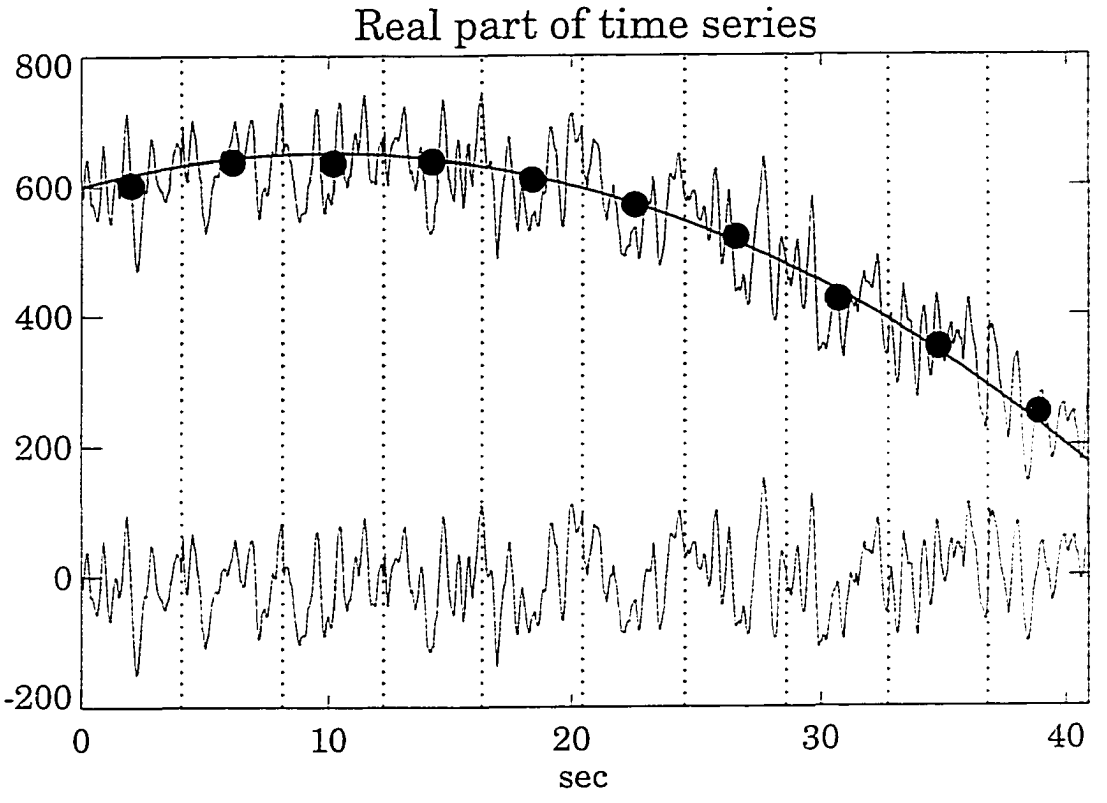


Figure A.1. Real part of a time-series for 10 blocks of 64 points each. Solid circles denote the mean values of each block, and vertical dotted lines delimit each block. The thick line represents the fitted second-degree polynomial to the circles. The ground-clutter-free signal is at the bottom, centered on zero.

A.1 Time-Domain subtraction

We perform a least-square second-degree polynomial fit to the mean values of each block ($\bar{R}_i$). Then this polynomial is sampled at the same points of the time-series, and subtracted from the time-series. In Figure A.1, the polynomial is represented by a thick line, and the ground-clutter-free time-series is centered at zero level. This procedure also works for a constant ground-clutter signal, where only the zeroth-order coefficient is different than zero.

The usage of this procedure is suitable for time-domain or for frequency-domain analysis. In both domains, time and frequency, once the "clutter" subtraction is performed in the time-series, the auto/cross correlations functions (ACFs/CCFs)

and self/cross spectrum functions (SSFs/CSFs) are computed for each block, respectively. Then these complex quantities are incoherently averaged to obtained the functions for that particular time-series, and the respective analysis follows (e.g., FCA in the time-domain, or FSA in the frequency-domain).

The main disadvantages of this procedure are: (1) the high amount of memory needed to do on-line processing (this is primarily due to the necessity to store n blocks before the time- or frequency-domain analysis is performed), and (2) for an off-line processing this procedure requires that the "raw" data be saved (i.e., voltages).

A.2 Frequency-Domain subtraction

We assume here that the Fourier transform of the complex voltages is given by:

$$s_f = a_f + c_f \tag{A.1}$$

where a_f represents the atmospheric and noise components, and c_f the ground-clutter component, at frequency-bin f. Furthermore, we assume that each A_f is a zero-mean Gaussian process, i.e., $< a_f >=< a_f^* >= 0$. c_f is a deterministic function and <> denotes incoherent averaging. Then, the power spectrum, after incoherent averaging, is given by

$$S_f =< s_f s_f^* >=< a_f a_f^* > - < c_f c_f^* > \tag{A.2}$$

(Woodman 1995, personal communication). Therefore the power spectrum of the atmospheric and noise contribution can be obtained by subtracting the ground-clutter deterministic power spectrum function, from the power spectrum of the original data. In Figure A.2, the original spectrum is represented with a thick line, the ground-clutter spectrum with pluses and the atmospheric, and the noise spectrum with diamonds.

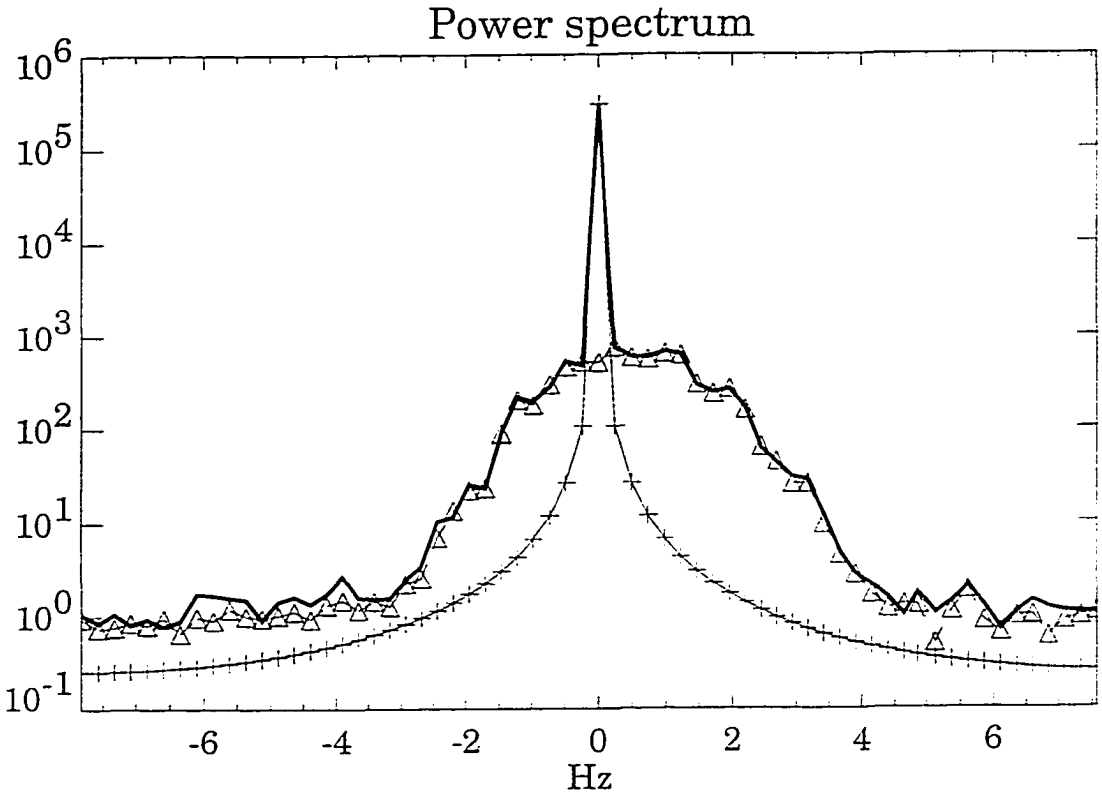


Figure A.2. Example of atmospheric power spectrum with ground-clutter and noise (thick line). Ground-clutter spectrum is denoted with pluses and the ground-clutter-free spectrum with triangles.

The deterministic spectrum $< c_f c_f^* >$ can be obtained by Fourier transforming the complex polynomial that was fitted to the complex mean values of each block. Then, the desired spectrum $< a_f a_f^* >$ is obtained from Equation A.2.

The main advantage of this procedure is that the ground-clutter removal is done after incoherent averaging, i.e., there is no need to keep the "raw" data in either an on-line or an off-line processing. Only $< s_f s_f^* >$ and the complex mean values of each block of the incoherent averaging process, are needed. This procedure, however, can not be applied in time-domain analysis. In addition, it requires a reasonable number of incoherent averages in order to better characterize the atmospheric spectrum.

For the case the ground-clutter signal is constant or near constant, there is

no need for fitting, and the only removal is done at frequency-bin $f = 0$.

A.3 Frequency-Domain interpolation

This last procedure is suitable only for frequency-domain analysis when ground-clutter signals are near constant (i.e., when most of the ground-clutter energy is concentrated at $f = 0$). Under these conditions the power at zero frequency, i.e., $< s_0 s_0^* >$, is replaced by an interpolated value. This value can be obtained from a simple linear interpolation using the two adjacent points, or from a cubic interpolation using two closest points at each side of $f = 0$.

APPENDIX B

SPECTRAL MOMENT ESTIMATION

The atmospheric signals obtained with VHF radars at any one height can be assumed to be Gaussian stationary processes with a Gaussian-shaped power spectrum (e.g. Woodman [1985]). In addition, to the atmospheric signals, the radar signals will be primarily composed of the following non-desirable signals: (1) ground clutter and (2) sky and receiver noise. The ground clutter is usually approximated to be a constant DC signal, although under certain circumstances it could have significant temporal variations. The sky and receiver noises, on the other hand, are well approximated by a white-noise process, i.e., a spectrum of constant value. Having said this, the power spectrum of the received signal can be assumed to be,

$$S(f) = \frac{P}{(2\pi\sigma_f^2)^{1/2}} \exp\left[-\frac{(f-f_d)^2}{2\sigma_f^2}\right] + N + D\delta(f) \qquad \text{(B.1)}$$

where f is the frequency in Hz and $\delta(f)$ is a delta function. Then, this spectrum is fully characterize by the value of five parameters: total atmospheric power (P), mean atmospheric frequency shift (f_d), atmospheric spectral width (σ_f), total noise (N) and the ground clutter power (D) (see Figure B.1).

In this section we give a brief description of the two spectral moments estimators used throughout this thesis: one used by the Aeronomy Lab of the National Oceanic and Atmospheric Administration (NOAA) and one that performs a Gaussian fitting.

B.1 NOAA's algorithm

This algorithm is also known as first moment (FM) and has been successfully compared, in terms of frequency shift estimation, to other algorithms, by May

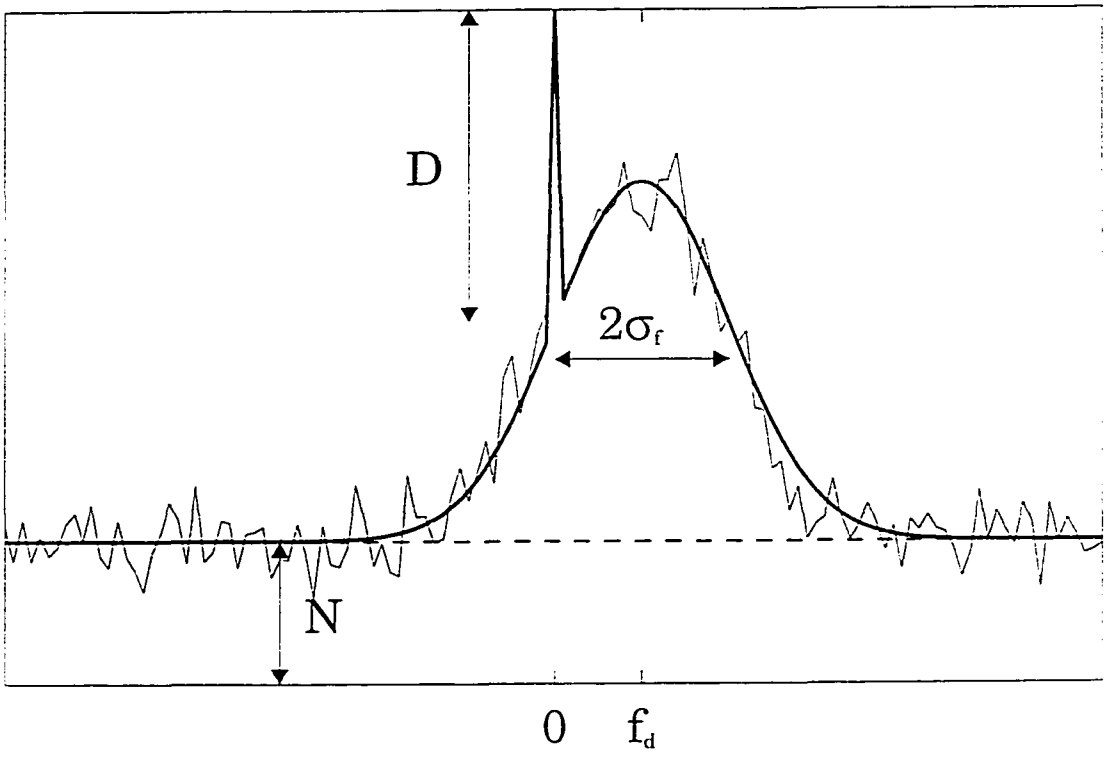


Figure B.1: An example of a frequency spectrum represented by Equation B.1.

and Strauch [1989]. Here we assumed that the ground clutter contribution has been already removed by one of those procedures described in Appendix A. Therefore the parameters to be estimated are: P, f_d, σ_f and N.

This FM algorithm involves several steps. First several Doppler spectra from each height range are averaged (Incoherent integration). Then the noise level (N) is objectively estimated using the technique of Hildebrand and Sekhon [1974]. A first guess for the position of the signal peak is obtained by finding the maximum spectral power density when the spectrum has been smoothed with a running mean of n points (usually 3). The moments (P, f_d and σ_f) are then calculated from the original (non-smoothed) spectrum over the interval surrounding this guess out to the first point where the spectrum power density falls below the noise level on either side of the guess. The noise level is subtracted from the spectral values over the signal interval before calculating the moments. Instead of total power (P), the parameter

that is of most common use is the signal-to-noise ratio (SNR = P/N).

The main advantages of this algorithm are that the moments are calculated quite fast, and the Doppler shift (first moment) estimates are very accurate. In contrast, however, the spectral width (second moment) is not accurately determined with this algorithm. More on this algorithm can be found in Strauch et al. [1984] and May and Strauch [1989].

We have applied this algorithm in Chapter 4 to compute the Doppler shifts of the synthesized beams used with the PSS technique. In Chapter 6, Doppler shifts of the Narrow and Wide vertically-directed beams were also computed with this algorithm.

B.2 Gaussian fitting algorithm

In this algorithm we fit the averaged Doppler spectrum to a known function that includes the theoretical shape of the spectrum (e.g. Equation B.1) and any instrumental function (e.g. a rectangular window). The concept is described in Figure B.2: For any single height, the outputs of the radar, i.e., the atmospheric signal, ground clutter, noise and the instrument function, are represented by the pairs f_i, S_i. These measured values are fitted by the theoretical values f_t, S_t in a least squares sense (Woodman [1996], personal communication).

In our case we have assumed that the theoretical spectrum is represented by Equation B.1. We have represented the instrument function by a "window", rectangular or Hanning (e.g. Press et al. [1992]). The theoretical pairs (f_t, S_t) are obtained by finding the inverse Fourier transform of the theoretical model, multiply it by the autocorrelation of the window used, and Fourier transforming back this product.

We can use this approach when the ground clutter has not been previously removed, or even when it has been removed. The latter approach will be in order to

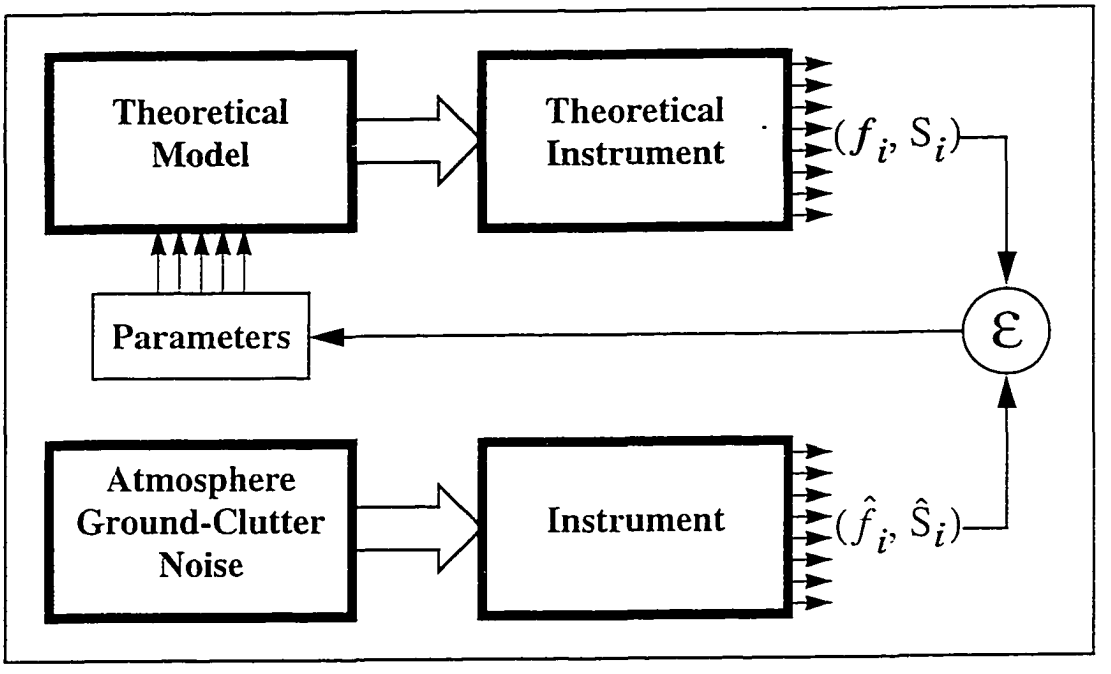


Figure B.2: Block diagram for a least squares parameter estimation.

removed residual "spikes" (not enough removal) or "notches" (too much removal). Furthermore, this approach can also be applied when ground clutter is not constant, but it is slightly changing. Or in the presence of other known, but undesirable signals (e.g. sea echoes, interference, etc.).

The main advantages of this type of algorithms are: (1) accurate measurements of the atmospheric spectral width, and (2) good characterization of the atmospheric spectrum when the "a priori" information is good. The main disadvantage of this algorithm is that requires higher estimation time than using the NOAA's algorithm.

In order to initialize the fitting procedure, we obtain the initial guesses from the estimates of NOAA's algorithm. In Chapter 4, we have implemented this algorithm in the full spectral analysis technique (FSA) and in the radar interferometric techniques (RIV, RIH).

APPENDIX C

PHASE CALIBRATION OF SPACED ANTENNA SYSTEMS

Some of the spaced antenna techniques for wind measurements are inherently dependent on any phase imbalance in the system (e.g. see Section 4.2 and Section 6.2.2). The imbalance is usually caused by a path length difference between the multiple receivers. One way to measure this imbalance is by feeding a common signal to all the receiving lines simultaneously (Kudeki [1994], personal communication). These receiving lines include cables, receivers and Analog/Digital converters. Once the signals are recorded, the phase difference should be obtained with a similar analysis used to measure the winds. This last step is necessary in order to avoid, mainly, problems associated with sign conventions.

In our experiments, we have done the phase calibration during 30 minutes, at the beginning and at the end of each experiment. Two types of common signals were used. First, the atmospheric signal of one of the receiving antennas (transmitter on) was equally split and fed to all the receiving lines. With this procedure, we were able to test the phase imbalance under different signal amplitudes. The second common signal was an amplified "sky noise" measured at one of the receiving antennas (transmitter off). In both cases, we measured almost the same phase differences.

In Table C.1, we show the mean and the standard deviation of the phase differences between receiving lines, for the 5-day experiment run in March 1997. Notice the good agreement of the values under both atmospheric and "sky noise" signals. The biggest difference of the mean values for each receiver pair is $\approx 1.1°$, that corresponds to an AOA error of $\approx 0.03°$.

Table C.1: Phase calibration values for the 5-day experiment run in March 1997.

	AB	AC	BC
Atmospheric signal			
Mean	$-82.2°$	$-109.4°$	$-26.1°$
Standard deviation	$1.2°$	$1.2°$	$1.7°$
Sky noise			
Mean	$-81.1°$	$-108.8°$	$-26.2°$
Standard deviation	$1.2°$	$1.1°$	$1.0°$

APPENDIX D

IMPLEMENTATION OF THE POST STATISTICS STEERING TECHNIQUE

In configurations using multiple-receiving antennas, different beam directions can be synthesized by phasing the received signals. One of such technique is post set beam steering (PBS) (Röttger and Ierkic [1985]). This technique synthesizes the beam by introducing a phase shift on the signal itself. Post statistics steering (PSS) (Kudeki and Woodman [1990]) improved this approach by introducing the phase shift on the signal statistics (CCF, CSF) rather than the signal itself, thus reducing computation and storage. In the following paragraphs, we describe how a three-antenna PSS approach was implemented in the frequency domain.

Using three receiving antennas, the synthesized spectrum ($\Phi_{\alpha\alpha}$) in a desired beam direction can be obtained from

$$\Phi_{\alpha\alpha} = \Phi_{aa}+\Phi_{bb}+\Phi_{cc}+2\mathrm{Re}\{\Phi_{bc}e^{j\Delta_{bc}}\}+2\mathrm{Re}\{\Phi_{ba}e^{j\Delta_{ba}}\}+2\mathrm{Re}\{\Phi_{ca}e^{j(\Delta_{ba}-\Delta_{bc})}\} \quad \text{(D.1)}$$

where Φ are the complex spectra functions of the original signals (auto and cross spectra), and Δ are the phase shifts.

In order to obtain the effective beam pointing direction, one needs to take into account the patterns of the transmitting and receiving antennas, the location of the antennas, and an atmospheric function. Using the antenna setup of Figure 4.1, and assuming isotropic scattering, the synthesized pattern at height z_0 is

$$g_s(x,y) = |g_T(x,y)[g_{Rb}(x,y) + g_{Ra}(x,y)e^{-j\Delta_{ab}} + g_{Rc}(x,y)e^{-j\Delta_{cb}}]| \quad \text{(D.2)}$$

where g_T and g_R are the patterns of the transmitting and receiving antennas. From Figure 4.1, baseline BC is aligned in the x direction, and baseline BA in the y direction.

The effective pointing directions are calculated from

$$\theta_x = \sin^{-1}\left(\frac{\overline{MB} - \overline{MC}}{d_{bc}}\right) \quad \text{(D.3)}$$

$$\theta_y = \sin^{-1}\left(\frac{\overline{MB} - \overline{MA}}{d_{ba}}\right) \quad \text{(D.4)}$$

where M is the peak of the synthesized pattern in distance, d_{ba} and d_{bc} are the distances between antennas A and B, and between antennas C and B, respectively. The distances of point M to the receiving antennas A, B, C are denoted by $\overline{MA}$, $\overline{MB}$, $\overline{MC}$, respectively. In addition, we have calculated the position of point M, by fitting a two-dimensional Gaussian function to the synthesized pattern, over the region surrounding the peak.

The three-dimensional velocity vector can be obtained with n synthesized directions by solving the following system of equations

$$\begin{bmatrix} \sin\theta_{x1} & \sin\theta_{y1} & \sqrt{1-\sin^2\theta_{x1}-\sin^2\theta_{y1}} \\ \vdots & \vdots & \vdots \\ \sin\theta_{xn} & \sin\theta_{yn} & \sqrt{1-\sin^2\theta_{xn}-\sin^2\theta_{yn}} \end{bmatrix} \begin{bmatrix} v_{0x} \\ v_{0y} \\ v_{0z} \end{bmatrix} = \begin{bmatrix} v_{r1} \\ \vdots \\ v_{rn} \end{bmatrix} \quad \text{(D.5)}$$

where v_{ri} is the radial velocity (positive, going away from the radar) of the i synthesized direction. We have solved this system of equations using a singular-value decomposition and a back substitution (Press et al. [1992], section 2.6).

We have synthesized nine different beam directions at each height. For example, the angles of the effective beam directions synthesized at 7.50 km are given in Table D.1. Similar values have been obtained at the other heights. Briefly, we have selected values around the original pointing direction trying to: (a) effectively point the furthest from vertical, and at the same time, (b) avoid the generation of significant sidelobes. By synthesizing the furthest possible off-vertical angles, we try to decrease the influences of the pointing angles uncertainties on the velocity estimation.

Table D.1: Effective beam pointing directions at 7.50 km (in degrees).

	1	2	3	4	5	6	7	8	9
θ_x	-0.36	-0.36	-0.60	0.11	0.08	-0.12	0.61	0.67	0.58
θ_y	-0.41	-0.06	1.17	-0.36	-0.50	1.43	-0.23	0.29	1.20

In Equation D.1, zero phase differences between receivers and identical receiver gains are assumed. The zero phase difference was accomplished by subtracting the non-zero phases obtained from the calibration procedure described in Appendix C. In order to have identical receiver gains, we have balanced the spectral functions using the noise levels of the "noisy" heights (between 12 and 15 km). This procedure was performed before the synthesized spectra were achieved. The procedure of getting the effective beam directions of the synthesized beams, can also be applied to a PBS approach.

APPENDIX E

AOA INTERPRETATION UNDER HORIZONTALLY ANISOTROPIC SCATTERING

Using Equation 21 from Holloway et al. [1997b] and the AOA contributions of Section 5.2.1, we arrive to the following normalized cross-correlation function.

$$\begin{aligned} c_{ij}(\tau) \approx \exp\Bigg\{ & -\frac{2(v'_{0x}\tau - \Delta\rho'_{xij}/2)^2}{\xi_x^2} - \frac{2(v'_{0y}\tau - \Delta\rho'_{yij}/2)^2}{\xi_y^2} \\ & -2(k_0\sigma_t\tau)^2 - \frac{1}{8}\left(\frac{v_{0z}\tau}{\sigma_r}\right)^2 + j\phi_{ij}(\tau)\Bigg\} \end{aligned} \tag{E.1}$$

where

$$\begin{aligned} \phi_{ij}(\tau) &\approx -2k_0 v_{0z}\tau - 2k_0 v'_{0x}\Theta'_{xij}\tau - 2k_0 v'_{0y}\Theta'_{yij}\tau \\ &\quad + k_0\Delta\rho'_{xij}\Theta'_{xij} + k_0\Delta\rho'_{yij}\Theta'_{yij} + k_0|\Delta\rho'_{ij}|\Lambda_{ij} \qquad \text{(E.2)} \\ \Theta'_{xij} &= 2(\Omega'_x + \Phi'_x + \Gamma'_{xij})\xi_x^{-2} \\ \Theta'_{yij} &= 2(\Omega'_y + \Phi'_y + \Gamma'_{yij})\xi_y^{-2} \\ \xi_x^2 &= 2a_h^{-2} + (2\rho'_{cx})^2 \\ \xi_y^2 &= 2a_h^{-2} + (2\rho'_{cy})^2 \end{aligned}$$

and ρ'_{cx} and ρ'_{cy} are the correlation lengths along the primed coordinate axes (x', y').

The irregularities are assumed to have the major/minor axes of their correlation ellipse aligned along the $x' - y'$ axes which are rotated by an unknown angle ψ with respect to the $x - y$ axes (the axes that define the transmitter and receiver locations).

Now, the primed coordinate system $(x' - y')$ is related to the $x - y$ system by the following transformation,

$$\boldsymbol{X}' = \boldsymbol{T}\boldsymbol{X} \tag{E.3}$$

where $\boldsymbol{X}' = (x', y')$, $\boldsymbol{X} = (x, y)$ and $\boldsymbol{T} = [(\cos\psi, \sin\psi), (-\sin\psi, \cos\psi)]$. Applying these transformation to all the primed variables (velocity, position and AOA contributions), the phase term given by Equation E.2 becomes,

$$\begin{aligned}\phi_{ij}(\tau) \approx\ & -2k_0 v_{0z}\tau - 2k_0 v_{0x}\Theta^a_{xij}\tau - 2k_0 v_{0y}\Theta^a_{yij}\tau \\ & + k_0\Delta\rho_{xij}\Theta^a_{xij} + k_0\Delta\rho_{yij}\Theta^a_{yij} + k_0|\Delta\boldsymbol{\rho}_{ij}|\Lambda_{ij}\end{aligned} \tag{E.4}$$

where $(\Theta^a_{xij}, \Theta^a_{yij})$ represents the observed AOA vector, and is given by,

$$\Theta^a_{xij} = \Theta_{xij}(\cos^2\psi + (\xi_x^2/\xi_y^2)\sin^2\psi) + \Theta_{yij}\cos\psi\sin\psi(\xi_y^2/\xi_x^2 - 1) \tag{E.5}$$

$$\Theta^a_{yij} = \Theta_{yij}(\cos^2\psi + (\xi_y^2/\xi_x^2)\sin^2\psi) + \Theta_{xij}\cos\psi\sin\psi(1 - \xi_x^2/\xi_y^2) \tag{E.6}$$

and

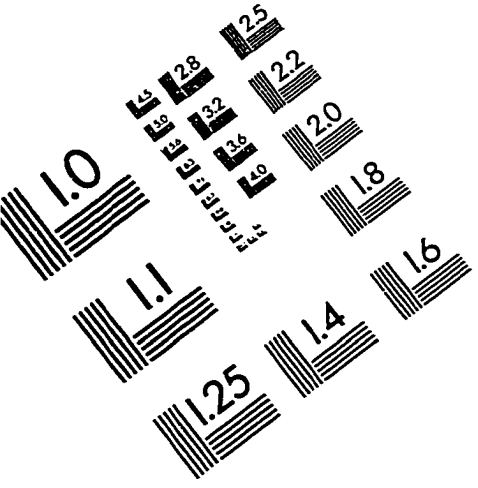
$$\Theta_{xij} = 2(\Omega_x + \Phi_x + \Gamma_{xij})\xi_x^{-2} \tag{E.7}$$

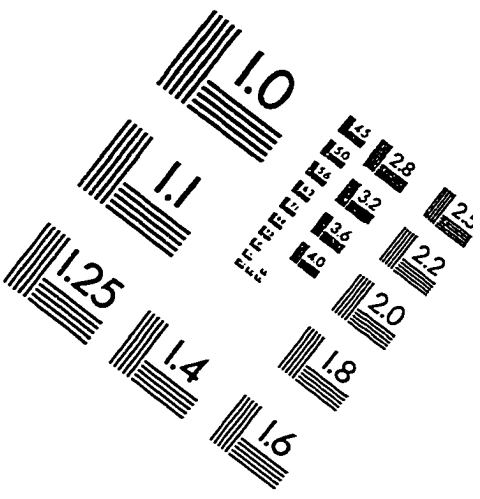
$$\Theta_{yij} = 2(\Omega_y + \Phi_y + \Gamma_{yij})\xi_y^{-2} \tag{E.8}$$

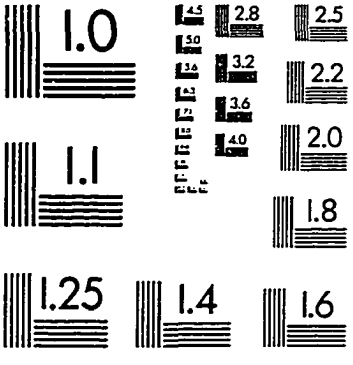
The influence of horizontal anisotropic irregularities on the statistical errors it is mainly through the correlation time (τ_c) and through the coherence term ($\mathrm{coh}_{ij}[0]$). Transforming Equation E.1 to the $x-y$ axes, the square of the coherence term become,

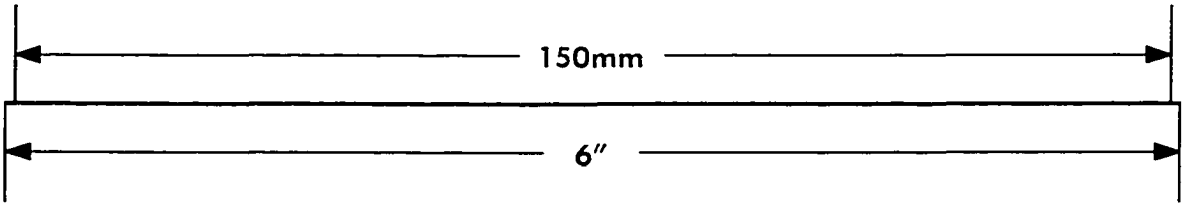
$$\begin{aligned}\mathrm{coh}^2_{ij}[0] = \exp\Big\{ & -[\Delta\rho^2_{xij}(\xi_x^{-2}\cos^2\psi + \xi_y^{-2}\sin^2\psi) + \Delta\rho^2_{yij}(\xi_y^{-2}\cos^2\psi + \xi_x^{-2}\sin^2\psi) \\ & + \Delta\rho_{xij}\Delta\rho_{yij}\cos\psi\sin\psi(\xi_x^{-2} - \xi_y^{-2})]/2\Big\}\end{aligned} \tag{E.9}$$

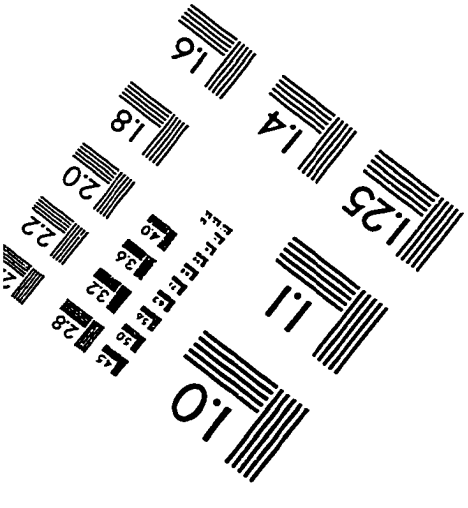
IMAGE EVALUATION TEST TARGET (QA-3)

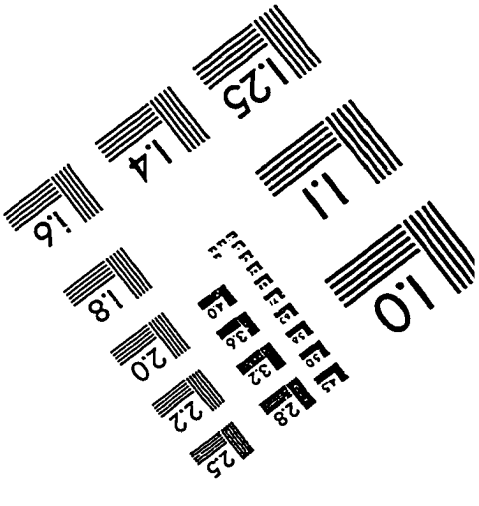












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Identity, Discourse, and Media Audiences: A Critical Ethnography of the Role of Visual Media in Religious Identity-Construction among U.S. Adolescents

by

Lynn Schofield Clark

B.A. Westminster College, 1986
M.A. United Theological Seminary, 1990

A Thesis
Submitted to the Faculty of the Graduate School
of the University of Colorado

in Partial Fulfillment of the Requirement for the degree of
Doctor of Philosophy
School of Journalism and Mass Communication
1998

UMI Number: 9827697

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

This thesis titled:
Identity, Discourse, and Media Audiences:
A Critical Ethnography of the Role of Visual Media in Religious Identity-Construction among U.S. Adolescents
written by Lynn Schofield Clark
has been approved for the
School of Journalism and Mass Communication

Chair of Committee
Stewart M. Hoover, Ph.D.

Committee Member
Janice Peck, Ph.D.

Committee Member
Andrew Calabrese, Ph.D.

Committee Member
Erika Doss, Ph.D.

Committee Member
Janet Jacobs, Ph.D.

Date 4/9/98

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Clark, Lynn Schofield (Ph.D., Media Studies, School of Journalism and Mass Communication)

Identity, Discourse, and Media Audiences: A Critical Ethnography of the Role of Visual Media in Religious Identity-Construction among U.S. Adolescents

Dissertation directed by Professor Stewart M. Hoover, Ph.D.

Employing a critical/cultural studies approach, this dissertation argues that identity-construction is best understood as the nexus of public discourses and individual subjectivities. To understand the role of media in identity-construction, this work analyzes both the themes of discourse that are available in mediated texts and echoed throughout the culture, and the various social, political, economic and other contexts that frame the individual adolescent's identity narratives and practices. The discourses of religion and their relation to the religious identity-construction of individual subjects provides the focus for the current analysis.

The study employed ethnographic interviews with 70 adolescents and their parents, 5 in-depth case studies of adolescents, 3 "peer-led" discussion groups (some of the adolescents involved in case studies were trained to lead focus groups without the primary researcher present), and 3 focus groups with parents of teens.

The dissertation argues that there are three distinctive elements of religious identity-construction among adolescents today. First is a *flattening of religious symbols.* Religious symbols are not necessarily seen by adolescents as authoritative and "fixed" due to their reference to formal

religious institutions but are rather approached as somewhat autonomous and, like other commodified symbols of the postmodern condition of late capitalism, they must be *made useful.* Second, analyzing the interpretive strategies teens brought to the popular television program *Touched by an Angel,* the dissertation finds that adolescents embrace a variety of publicy-available discourses of religion which are not solely attributable to race, class, gender, and religious affiliation. Thus the dissertation affirms the rise in personal autonomy or the privatization of religion and the subsequent importance of the mediated realm (as opposed to solely the realm of religious institutions) in determining religious identities. Third, while affirming Stuart Hall's interpretive taxonomy of dominant, negotiated, and oppositional readings, the dissertation demonstrates a fourth interpretive approach, a *regeneration* that draws upon a dominant or negotiated reading of a text and is based on a viewer's position with reference to the text, yet also subtly informs the individual's larger system of beliefs, thus resulting in a subtly changed belief system.

Dedication

for Jon

who got me involved with teens in the first place
and supported me all the way through

Acknowledgments

The research and analysis in this dissertation was made possible through a generous Dissertation Fellowship from the Louisville Institute. Research conducted with the "Symbolism, Media and the Lifecourse" project under the direction of Stewart Hoover, funded by the Lilly Endowment, Inc., also afforded significant opportunities for research and consultation, and I am deeply grateful for the opportunity to serve as Associate Investigator with this project.

My first thanks must go to Stewart Hoover, who has gone well beyond the call of academic duty in his role as my advisor and mentor. Stewart is not only brilliant but also a joy to work with, and I feel very honored for the many opportunities he has afforded me in my professional path. His discipline and vision serve as a constant inspiration to me.

Thanks to those who provided detailed feedback and comments on early drafts of dissertation chapters and on the final project: Janice Peck, Andrew Calabrese, Erika Doss, Janet Jacobs, Alf Linderman, Richa Nagar, and Dennis McGilvray.

Thanks to the fellow members of the International Study Commission on Media, Religion, and Culture; those participating in the consultation on reception methodology in May, 1997 in Uppsala, Sweden; those participating in the consultation of Louisville Institute Dissertation Fellows; those attending presentations of the research at various academic conferences; and to those participating in the UCSB generations seminar on methodology - all of whom provided helpful feedback on this project.

I would also like to thank others who served as consultants at various times in the formulation of methodological and theoretical considerations of this project, including: Wade Clark Roof, Nancy Ammerman, Brenda Brasher, Judith Stacey, Carolyn Marvin, Don Roberts, Robert Kubey, David Morgan, Ron Grimes, Robert Wuthnow, Carol Lytch, Diane Winston, Joy Charlton, William McKinney, Carl Dudley, Knut Lundby, Klaus Bruhn Jensen, Birgitta Hoijer, Ingunn Hagen, Elizabeth Bird, Robert White, Michele Rosenthal, Jane Iwamura, James Lull, Kristine Greenaway, Randy Naylor, John Peterson, Lois Ford Long, Allen Angell, Sheila George, Shirley Whipple Struchen, Adan Medrano, Shu-Ling Berggreen, Diane Alters, Lee Hood, Joe Champ, David Shaw, Jan Fernback,Tom McQuaide, Sam Smith, Greg Stene, and Mark and Karen Borchert.

I would like to thank Ken Bedell, Dennis Benson, Gregg Hartung, Karen MacDonald Johnson and the "Power Women." Finally, I would like to acknowledge all of those supporters of this project who served as attentive listeners long after it might have been humanly expected of them: Carrie and Jim Bohrer, Meg Weisbeck, Jennifer and Michael Olin-Hitt, Katrina Sarsen and Nick Morris, the "Renaissance Babes," Karen Hoover, Gun Linderman, Patty Gassaway, the Ramey family, the Bucey family, the Pardridge family, Mother and Dad Clark, Robert Schofield, my parents Sandy and Jerry Schofield, and most of all my husband and life partner, Jonathan Clark.

Finally, I want to thank the teenagers and their families who shared so much of their lives with me in an effort to help me and my readers understand their experiences. Without you, there would be no story to tell.

CONTENTS

Signature page	ii
Abstract	iii
Dedication	v
Acknowledgments	vi
Chapter One: Introduction: Exploring Teens as Audiences	1
Chapter Two: The 'Self' in Meaning-Making: A Critical/Cultural Approach to Identity-Construction at the Site of the Audience	35
Chapter Three: The Method of Critical Ethnography: Linking Reflexive Empirical Research with Critical Theory	89
Chapter Four: Adolescents as Selves, Adolescents as Audiences	140
Chapter Five: Case Study: Elizabeth as a "Universal Singular"	177
Chapter Six: Parents, Teens and the Media	212
Chapter Seven: Religion and the Media: Teens and the Public Discourses of Religion	250
Chapter Eight: Conclusion	301
References	321
Appendix	344

Chapter One

Introduction: Exploring Teens as Audiences

"Why don't you go to church anymore?," I once asked Tom, a 14-year-old African American whose mother was a neighbor and friend of mine. He shrugged, looking down at his Keds. "They don't speak my language," he said. "And what's your language?," I pressed. "Music," he said. When I reflect on the current research project to be described here, it strikes me that Tom's story is quite pertinent. Many religious institutions in the U.S. are in decline, in part, I believe, due to the failure of these institutions to "speak the language" of the multicultural, expressive, tolerant, and largely mediated culture in which they exist. Yet at the same time, religion - or at least, the "God language" of religion - still holds an important place in the lives of many Americans, a fact overlooked in most media audience research that attempts to account for the role of music and other media in the meaning-making audience practices of young people. Ironically, it seems that scholars in both the fields of religion and media do not "speak" each other's language very well at all.

In the past decade, however, a host of scholars in American religion have asserted the need to look outside religious institutions to such places as the media sphere for clues as to the evolution and current status of what Lippy refers to as "popular religiosity" (Lippy, 1994; Roof, 1993, 1997; Wuthnow, 1988, 1992; Albanese, 1993). Some scholars of the media have begun to address the ways in which religion has been misunderstood and

overlooked in both news and entertainment media, as well (Wright, 1997; Hoover, forthcoming). Yet even as gender, race, class, geographic and ethnic identity have come to be key concerns within the field of media studies, few have explored the intersection of these various aspects of identity with religion.[1] On the whole, most scholars in the media fail to recognize the polyvalence of religious experience in the U.S. and the subsequent and varied role religious identity plays in the meaning-making processes of media reception. On a similar note, scholars in religion overlook the polysemic nature of media texts and the complex cultural, political and economic roles of media industries within the public sphere of the U.S. and in the private lives of its constituents. Often, the media are seen simply as "vehicles" through which religious organizations communicate (or fail to communicate) public aspects of their identity.[2] Studies of adolescence in both fields are a case in point. In media studies, religion is assumed to play a marginal to nonexistent role in the lives of young people in the U.S., despite evidence to the contrary.[3] Likewise in the studies of religion and youth, the media are either conceived of as a dangerous "force" to be combatted, or they are overlooked altogether.

By exploring the negotiative practices of meaning-making that occur as differently-positioned young persons and their friends encounter

[1] There are many examples of studies in media and identity which leave out any discussion of religion (see, e.g., Hall & du Gay, 1996; Morley & Robins 1995; Radway, 1984; Press, 1992; Craig,1992). Yet some have explored these negotiations and their role in audience reception practices (see, e.g., Naficy, 1993; Gillespie, 1993; Hebdige, 1976).

[2] An example of the use of this term to describe the role of the media in religious research is found in Wuthnow (1994, p. 8).

[3] In a national survey, the George H. Gallup International Institute (1992, p. 23) reports that 95 percent of adolescents claimed a belief in God. Another survey found that more teens participated in church youth groups than in any other kind of organized activities such as vocational clubs, student government, or community youth organizations (Bachman, Johnston, & O'Malley, 1993, pp. 223-224).

and make meaning of the culturally constructed texts of the media, this dissertation addresses how young people with various religious backgrounds construct their religious identity as they interact with the religious discourses available in the public and mediated realm, with their parents who may or may not have a vested interest in seeing their religious traditions respected and continued, and with their peers who often do not share their religious background or beliefs. The work is a critical or "documentary" ethnography, in which the goal is similar to that expressed by anthropologist Sherry Ortner (1993): "to enter relatively small life-worlds and examine how large-scale social forces work themselves out in everyday life (p. 413)." Through analyses of several case studies, extensive interviews, and focus groups, I aim to shed light on questions of generational difference and cultural change with regard to issues of religion and of media, particularly as reflected in attitudes toward media, religious pluralism, multiculturalism, and tolerance. I hope to provide new theoretical and methodological insights, employing neo-Marxist poststructuralist and critical postmodernist concepts to challenge the current developmental and behavioristic assumptions governing the study of U.S. teens, their identity-construction processes, and their reception of media. Thus I intend to challenge essentialist notions of a "religious" audience and monolithic assumptions of the mediated realm by demonstrating how young people negotiate their identities within the boundaries of race, class, religion, gender, and position in relation to media.

In the next few pages, I hope to lodge this dissertation in the scholarly bodies of literature to which it refers and introduce the chapters

that will follow. By addressing the ways in which each of the elements of the current study are relevant and interrelated, I will demonstrate the assumptions that have guided previous works in these areas while also offering a preliminary explanation of how I plan to either employ or challenge these approaches in what follows.

Why Media?

A few decades ago, communication scholar James Carey (1975) made the intriguing observation that research into mass media audiences had been limited by the implicit *transmission* metaphor of communication. Arguing that the one-way "sender to receiver" model of communication oversimplifies the process, Carey introduces a *ritual* metaphor of communication in order to emphasize the ways in which practices surrounding mass media - such as the daily reading of a newspaper - satisfy more than simply a consciously sought need for information. Carey challenges the assumption that what the sender of a communication message *intends* is directly related to what the receiver *interprets.* This, of course, is not a new idea, and in fact had resonance with the poststructuralist "interpretive turn" in literary theory which problematized a fixed notion of meaning inherently tied to (and transported to the receiver by) a text. Prior to the beginning of this century, the function of language in culture was believed to be just this transparent and fixed, yet this changed with the introduction of semiotics by Swiss linguist Ferdinand de Saussure (1917). In his study of the relationship between the signfier (or symbol) and that which it signifies, Saussure argues that meaning is produced as objects are differentiated

from what they are not, e.g., as what *is* is distinguished from its "other." There is no "essential" relationship between a certain object and the symbol that represents it, therefore; the relations between elements of the symbolic system provide the meaningfulness of the system as a whole and of each of its parts. By extension, all elements of culture, including those mediated, are best understood within a context of a larger system of values. As Turner (1990) points out, the consequences of this interpretive turn is that

> Reality is made relative, while the power of constructing 'the real' is attributed to the mechanisms of language within the culture. Meaning is revealed to be culturally grounded -even culturally specific...Culture, as the site where meaning is generated and experienced, becomes a determining, productive field through which social realities are constructed, experienced, and interpreted (pp. 14-15).

The subject of study for those interested in the role(s) of mass communication within society, therefore, has of necessity been widened, particularly to address the interrelation of symbolic forms, their institutional creation, and their interpretation. As Carey (1975) suggests, communication is "a symbolic process whereby *reality* is produced, maintained, repaired, and transformed [italics mine] (p. 23)." Yet how do the media produce, maintain, repair and transform reality? Stuart Hall argues that this occurs unintentionally as media producers, like audience members, draw definitions of the situation from within the wider cultural and political economic structures of which they are a part. This process results in a *specific* representation of reality rather than a neutral reflection of ideas, as Hall (1982) argues:

> it implies the active work of selecting and presenting, of structuring and shaping; not merely the transmitting of an already-existing meaning, but the more active labor of *making things mean.* It was a

> practice, a production, of meaning: what subsequently came to be defined as a 'signifying practice (p. 64).'

The media, then, become central to the maintenance - or the disruption - of a culture's frameworks of meaning. My understanding of this process is rooted in Hall's argument: that in order to engage an audience and to communicate with them in a meaningful way, the media are required - whether they do so intentionally or not - to operate within a general framework of consensus for definitions of 'reality.' By drawing attention to the ways in which an audience is also a commodity to be gathered and then "sold" to advertisers, Hall's writings, influenced by Marx and later Gramsci, foreground the economic aspects of media and its relation to "popular" (or working class) culture (Hall, 1981). When considered in its historical location, Hall's writings were important in the subsequent development of an important political-economic tradition of research into media institutions. The media have been portrayed in these writings as reinforcing the predispositions of the members of the dominant (bourgeois) culture (See, e.g., Golding & Murdock, 1981). While this tradition has revealed important facets of the multinational monopolies that now organize our media industries, in its structural focus there has been a tendency to *assume* rather than *analyze* the agency of the structures of media organizations and the messages they deliver, hence overlooking issues of individual agency in the meaning-making processes of the audience. Indeed, some researchers from this tradition have argued that exploring individual meaning is an irrelevant question, accusing the cultural studies approach that has foregrounded these issues as "uncritical" (McGuigan, 1992, p. 244) and even "baleful" (Garnham, 1995,

p. 69). The current "debate" in media studies finds political economy pitted against cultural studies around exactly this issue of individual agency. Do audiences have agency in their ability to make meaning from mediated texts? Yes, assert many audience researchers, pointing to the unique ways in which people interpret what they see, hear, or read (Radway, 1986; Lull, 1990). Yet ultimately, Marxist- and British cultural studies-inspired researchers insist, social, political, and economic systems are determinative of meaning-making practices (Turner, 1990; Morris, 1990; McGuigan, 1992; Stabile, 1995). To which most current critical/cultural studies reseachers, myself included, would reply yes, of course (Jensen, 1995; Press, 1991). Only reductionist Marxism suggests that all meaning is created within and totally determined by ideology in a closed system; a more nuanced reading of Marx (or at least, neo-Marxism), as will be detailed below, suggests that social change can only occur when agency and individual consciousness are possible, even if unlikely.

The interesting question for current critical/cultural studies of media, then, is: how may we describe this interaction between sites of determination in the text, media industries, and audiences? It is this question, then, that guides the methodology and theory of this dissertation, as will be explicated more fully in the following chapters.

Why discourse?

As noted above, I am approaching the communication media as a site in which reality is publicly produced and reproduced, both in its production and its reception. The works of Gramsci (1971) have been influential in the cultural studies approach to this issue, and in my own

work as well. Drawing upon Marx's ideas of the relationship between the ruling class and the dominant ideology, Gramsci developed the concept of *hegemony* as the process by which the dominant group secures the support of others through consent rather than force. These 'ruling ideas' are accepted as the 'common sense' of a culture:

> Every social stratum has its own 'common sense' and its own 'good sense,' which are basically the most widespread conception of life and of men. Every philosophical current leaves behind a sedimentation of 'common sense': this is the document of its historical effectiveness. Common sense is not something rigid and immobile, but is continually transforming itself with scientific ideas and with philosophical opinions which have entered ordinary life...Common sense creates the folklore of the future, that is as a relatively rigid phase of popular knowledge at a given place and time (p. 326).

Gramsci's writings have made many contributions to media studies that we will not deal with at this point. However, while he is interested in why people allow themselves to live in "false consciousness," his writings also direct us to question why culture has a certain configuration and how certain ways of approaching the world come to be accepted as *the* natural way, or "reality." Stuart Hall (1980) writes of this challenge:

> First, how did a dominant discourse warrant itself as <u>the</u> account, and sustain a limit, ban or proscription over alternative or competing definitions? Second, how did the institutions which were responsible for describing and explaining the events of the worlds - in modern societies, the mass media, par excellence - succeed in maintaining a preferred or delimited range of meanings in the dominant systems of communication? How was this active work of privileging or giving preference practically accomplished? (p. 67)

Hall uses the concept of discourse to describe the form in which certain views are expressed as hegemonic. Yet as Nagar (1995) points out, "there is no absolute or conclusive hegemony - for hegemony describes the social processes by which a particular dominance is challenged and new forces

are articulated (p. 44)." As I will describe in chapter two, I define the concept of discourse with reference to both of these positions, as the culturally specific frameworks within which all narratives are constructed, whether they be mass mediated or emerging in interpersonal communication. They are not formally articulated, nor are they specifically created by certain organizations or individuals; they are not the sum of the conversations themselves, nor are they straightforward expressions of a fixed ideological position. The concept of discourse I will employ simultaneously point to the ways in which narratives tend to be patterned so as to emphasize and naturalize certain ideological commitments, while also noting the ephemeral nature of any such articulation. We can name the "themes" of discourse, therefore, by defining the key components of these frameworks around which certain patterns of naturalized meaning tend to cohere.

The concept of discourse is helpful in the current analysis, therefore, as it links individual narratives to larger patterns of public frameworks of meaning. It assumes that in order to communicate meaningfully, whether in interpersonal conversation or through the symbols of visual media, we all must rely upon certain consensually-accepted ways of seeing and comprehending. These frameworks of discourse are not neutral, but must be understood within existing systems of power as it operates through the cultural categories of race, gender, class, etc. "Meaning" happens within the limits of discourse; discourse sets the boundaries for how things can "mean." This is why it is important to understand discourse not simply as what a single media text "says," but as that which contributes toward framing both the meaning of the text and

the varied possible interpretations audiences can take from it. How meaning is made is therefore dependent upon how one is situated in relation to the culture's available discourses.

The in-depth analysis of individual narratives examined here, therefore, is not idiosyncratic, nor is it solipsistic. By exploring patterns in the narratives that emerge regarding the topics of media and religion, we can come to understand what is accepted or "hegemonic," what is contested, and what is negotiated. This then provides insight not only into the individual's meaning-making processes, but into how certain views of society are held in place as well as how some of these views may have shifted - or may be expected to shift - over time.

Why identity?

My argument is that we are able to contextualize the meaning-making processes of individuals within larger questions of discourse by exploring the narratives of identity. I take identity to be not simply the sum of labels that describe an individual's demographics, nor the descriptor which refers to the "core" or "essence" of one's self, but as an elusive analytical category which may be helpfully employed to explore the interrelationship of the individual and his or her social and material environment. Identity, as it is constructed in narrative, foregrounds the individual's perspective on that relationship, yet the individual must draw upon existing categories to either describe or differentiate. Thus identity is constituted in the symbolic marking of boundaries between self and other. Further, while the identity narrative itself may be seen as temporary and constructed for the specific audience listening at the time,

the frameworks of discourse upon which it draws are lodged culturally and historically. When a person constructs a narrative of identity, therefore, he or she is not simply speaking the subject position of the discourse as he or she is 'hailed' in it, but is actively claiming the narrative and investing in it. The narrative becomes an *articulation* of identity, to use Hall's (1996) term.

The relationship between the individual subject and discourse is far from a settled issue, however. In 1979, Hirst made his devastating critique of Althusser's notion of the subject as articulated in and through discourse. Hirst notes that in order for the subject to be 'hailed' by the discourse, he or she would need the capacity to recognize him or herself as a subject before that subjectivity was constituted (cited in Hall, 1996, pp. 7-8). Yet Foucault, maintaining Althusser's focus on power relations as constitutive of the discourse which then "speaks" the self, argues that the self is simultaneously *subjected* and *subjectified:* there is a conscious being in a material body that performs the production of self through the rules and norms of discourse (McNay, 1994, p. 87).[4] This is central, then, to Hall's (1996) definition of identity as

> the meeting point, the point of suture, between on the one hand the discourses and practices which attempt to 'interpellate,' speak to us or hail us into place as the social subjects of particular discourses, and on the other hand, the processes which produce subjectivities, which construct us as subjects which can be 'spoken.' Identities are thus points of temporary attachment to the subject positions which discursive practices construct for us (pp. 5-6).

Ironically, much of the explorations of identity in the writings of cultural studies - Hall's work included - devote little attention to the examination of the *practices* whereby subjects constitute themselves within discourse

[4] See, e.g., Foucault, 1988, 1987. I am indebted to Hall (1996) for the description of Foucault's contribution to the concept of the subject as outlined here.

(See, e.g., Hall & duGay, 1996; Morley & Robins, 1995). The current project, therefore, attempts to address this lack, linking Hall's concept of *articulation* (the claiming of the subjective position within discourse) with Bourdieu's (1977) notion of the *habitus* (the ways individuals exercise agency within the constraints placed upon them by the political, economic, or social structures in which they live). In both Bourdieu's and Hall's writings, power is a constraining force within culture. To best understand its workings in identity-construction, therefore, I argue that we must examine the practices of how adolescent subjects both speak and are spoken by the discourses of religion that are available in the popular mediated realm.

The foregrounding of discourse and other forms of social power within considerations of identity have only recently entered the purview of reflections on identity within the field of media studies, however, as behaviorist and psychological approaches have dominated media studies of identity in the U.S. for nearly half a century. Media research on identity has most frequently been addressed within questions of causality: do the media have this or that *effect* on one's behavior or cognitive development? (See, e.g., Pearl, Bouthilet, & Lazar, 1982) Identity in these paradigms has been understood in a rather fixed way as those components which, in their sum, might describe a "self." By categorizing various aspects of identification such as race, gender, or class as "variables," the research has assumed both that these categories may be understood discretely (and as abstracted from "lived" experience), and that the individual who embodies them will act in a way consistent with the expectations attached to the various categories. This is troublesome not

only in the U.S. positivist tradition but also in a European tendency to privilege class as the central category of analysis rather than one that may work in concert with others. As Somers (1992) writes in her insightful rethinking of the birth of the English working class:

> To understand action, however, why should we assume that an individual or collectivity has any particular set of interests simply because one aspect of their identity fits into one social category?...To let 'class' stand for a determinative experience is to presume that which has not been empirically demonstrated - namely, that identities are foundationally constituted by categorization in the division of labor within a unified socioeconomic totality we call "society (p. 607)."

I find Somers' work helpful both in pointing to the myriad social relationships in which people are embedded and to their inevitable interrelation. However, I do not agree with her argument that identity should *replace* the concept of interest; instead, I believe it is helpful to consider, as Nagar (1995) notes, that interests and identities are intimately linked:

> Identities are self-understandings held by individuals and groups on the basis of their everyday social relationships and experience. The interests that motivate people to act in certain contexts are derived from their particular identities and such actions often reinforce those specific identities, or give birth to new ones (p. 44).

I believe that both Somers' and Nagar's approaches help to challenge the assumptions of a self comprised of fixed, causal categories of identification. Yet the questions remain: in any empirical investigation, are we always limited in the extent to which we can address and describe the *dynamic and lived* aspects of identity? Perhaps in order to operationalize its research questions, all empirical research requires an abstraction from "real time," the results of which can at best be considered a snapshot which does not reflect the *processural nature* of identity-construction. In

the next section, I will attempt to further develop the concept of identity and its methodological investigation.

Why critical ethnographic approaches?

Both quantative and qualitative approaches to mass media research begin with the observation of real-world phenomena, although quantitative researchers tend to treat what is observed as a set of variables with discrete properties which lend themselves to study as they are operationalized along measurable lines. This renders the approach quite appropriate for certain questions of audience research such as the number of hours spent viewing visual media, or the number of persons who agree with certain statements of public opinion. Qualitative researchers, however, argue that problems arise when quantitative research is employed to determine causality regarding the influence of various social factors on the individual behavior or beliefs expressed.

The design of this qualitative research project grows out of developments in feminist methodology in the past few years which have emphasized the dialogical aspects of knowledge production as well as the exposing of power relations in society. As Van Zoonen (1994) notes, "what feminist research has to add to interpretive research strategies is a notion of power, an acknowledgment of the structural inequalities involved in and coming out of the process of meaning making (p. 134)." This approach stresses the fact that knowledge emerges from within the context of the interview, as researcher and research participant interact. Feminist scholars argue that there is no completely impartial and neutral position from which a researcher might observe and record data from the research

participants' lives, pointing out the distortions that arise from such an approach. They argue that the researcher's perspective is always informed by her or his gender, class, race, and culture, among other factors. The way the researcher is "read" by the research participant is another fact influencing the knowledge generated, and thus narratives constructed in interviews must always be understood contextually (See, e.g., Personal Narratives Group, 1989). They are constructed for communication within the unique situation of an interview exchange, and as such they are tentative statements, consistent with the participant's self-understanding in the context of his or her relationship with the interviewer. Thus researchers are called upon to reflect on their own lives at the moments of the interview and its analysis, thereby acknowledging that research participants are active, self-conscious narrators articulating identity within a wider, culturally defined discourse. This reflexivity also draws attention to the fact that the researcher, like those who are researched, exists within a particular constellation of social identities which condition access to social and material power and shape the interview experience because it frames the way researcher and research participant view one another. Our participation in networks of social identity is what makes us understandable to one another and the patterns of these interactions then suggest categories of analysis.

I use the term "ethnography" to describe my work for two reasons. First, the term is used to mark the alignment of my work with the traditions of cultural anthropology, feminist studies, the humanities, and interpretive sociology as opposed to an alliance strictly with the social sciences. Second, while I admit that my research does not satisfy the

anthropological definition of ethnography which requires daily participant observation, attention to context, and contact with research participants over a period of a year or more, the use of the term by media researchers generally suggests a depth of interaction with research participants that moves beyond the single-interview approach common to social scientific qualitative research in media studies.[5]

Both the attention to reflexivity and my theoretical grounding in critical neo-Marxism further suggest the use of the term "critical" as a modifier. As I will discuss in chapter three, the methodology employed allowed me to approach the research not as an established set of hypotheses to be tested, but as a *process* of learning and of continuous interaction between theory, data collection, and analysis which reshaped and honed the project over time.

Why Teens?

While scholarship on the relation of media and identity-construction among *adults* covers a wide range of subjects and ideological positions as I will discuss in chapter two, scholarship specifically focused on *adolescents* has tended to occur within certain public discourses of concern for youth of all ages, growing out of our assumptions of childhood as a "protected space" within the culture. For instance, parents and educators worry that their children watch too much television, spend too much time with video games or online, or are negatively influenced by song lyrics. Scholars, media critics, and policymakers respond with research demonstrating the dangers of these media, detailing how young

[5] For a discussion of the use of the term "ethnography" in media studies, see e.g. Van Zoonen, 1994.

people imitate undesirable behavior, idealize the consumptive values of capitalism, or accept a trivialized view of life based on what they see and hear (Bandura, Ross, & Ross, 1963; Postman, 1985; Atkin, 1990). While some of these concerns are expressed in the context of possible negative effects on adults, the situation with our young people seems much more urgent to parents and other leaders.

To a large extent, these concerns are in concert with the assumptions of developmental and cognitive psychology that pervade both scholarly work and popular knowledge about youth. As I will argue in chapter four, the developmental and cognitive psychological works of Piaget, Kohlberg, and Erickson have been influential in studies of youth to such a degree that approaches from other fields such as sociology, anthropology, or history are largely overlooked. In its interdisciplinarity, this project therefore aims to begin with an analysis of the context in which young persons make meaning from their particular cultural and historical experiences. It is thus intentionally focused on the present moment, exploring how teens make sense of their encounters with media and how these meanings are then incorporated (or not) into their self-reflexive projects through narrative.

One of the main reasons this study of U.S. adolescents is justified is that audience research on teens in the U.S. has only recently turned toward the ethnographic, and thus there is a dearth of research that explores how teens discuss their meaning-making practices. Second, when ethnographic studies of teens have been done, they have largely paid attention to either their role within youth subculture(s) or within families, but have not looked at both. Thus this study makes an attempt at

a more complete contextualization in its consideration of the various ways and places in which teens talk about what is meaningful to them. Third, teens are just learning to express themselves through narrative and thus while explanations are often halting or at times less coherent that those with adults, teens are also less aware of social sanctions which frequently inhibit the conversations of adults. This results in interesting combinations of negotiation within various discourses that are particularly intriguing. A final reason justifying this study is that today's adolescents are largely the children of American "baby boomers." By studying them, therefore, we may consider how this noteworthy generation reproduces its ideas of meaning and value through its children.

There are also several social trends which suggest that a study of teens might provide insight not only into how these teens will live when they are adults, but into what the world in which they live will look like and how this generation might see a continuance of certain trends toward multiculturalism, relativism, tolerance, increased individualism, and a fractured collective (Giddens, 1991; Lyotard, 1979; Ferguson, Gever, Minh-ha, & West, 1990). These social trends will be described in the final section of this introduction, as the study is contextualized demographically and historically.

Why visual media?

In the burgeoning fields of cultural studies and media reception studies, there have been several important research projects detailing the role of popular music in teen peer groups (Frith, 1981; Robinson, Buck, &

Cuthbert, 1991; Roe, 1983; Fornas, Lindberg, & Sernhede, 1995; Mitchell, 1996). Certainly, music tastes and preferences are important in the everyday lives and practices of identity among teens. Music is viewed as a key medium for teen individual and subcultural group identification, particularly as television viewing drops from the period of young childhood to older adolescence. However, teens still devote an average of 22 hours a week to television viewing - some of which includes viewing movies on their VCR or on cable television (Johnson, 1993). In another study, nearly three-quarters of the teens surveyed reported that they watched television or movies almost every day (Bachman, Johnson, & O'Malley, 1993, p. 221). While music-listening may occur in various environments in which the teen exists, viewing of television and movies largely occurs in the contexts of teens' homes - whether their own or their friends'. Hence, these are "family" media. One study of family media practices found, for example, that more than 40 percent of U.S. families reported that the whole household watched movies or television programs together very frequently, while only 8 percent said they never did (Ethan Allen, Inc., 1986, p. 72). An analysis of the visual media therefore may extend the discussions of teen identity-formation to a fuller exploration of the ways in which teens, their families, and their peers interrelate in the formation of media practices and meaning-making.

Another reason for concentrating on visual media in this study is that television and movies, in particular, are frequently a scapegoat for the problems of society. Our cultural anxieties about the need to protect "our children," coupled with our sense that images are persuasive and "powerful," are often expressed in terms of nervousness surrounding

consumption of the visual media. In part because they are so ubiquitous, common sense suggests that they are key to the myriad of problems we bemoan in society, from prejudice and racism, sexism, or classism, to violence and abuse. If only we could "unplug the plug-in drug," so the argument goes, our society would be simpler, kinder, and more just (Winn, 1985).

Or would it? Although it is politically expedient to say that it would, most of us know that in reality, the problems of society go much deeper than that. I believe that any informed cultural or social analysis which is to contribute to media policy concerning young people (whether global or local) should begin with a detailed analysis of the role of media in young peoples' lives and meaning-making processes. While this dissertation will not explore current U.S. policies regarding media and youth, it is situated as an alternative to current research which does purport to address policy concerns by attempting to locate causal direction in the media/youth relation. It is part of my task here to simply challenge those assumptions and approaches to the study of media and young people, which will be detailed more fully in chapters four through seven.

One of the intriguing aspects of a study on teens and media use in the household context which further justifies this study is the set of ambivalences that arise concerning the visual media. In Bachman, Johnston, and O'Malley's (1993) survey of U.S. highschoolers, for instance, most teens reported that their parents never restrict the amount of time they spend watching television, and not surprisingly, they are happy with this arrangement (p. 226, cited in Chadwick & Heaton, p. 87). Another survey found that more than three-quarters of parents of teens do not

believe that their children are spending too much time watching television, and more than 40 percent of the parents think that television does their children more good than harm (Peter D. Hart Research Associates, 1996, p. 3). However, half of the parents also admitted that they believe their teenage children watch inappropriate programs (p. 4). I believe that this lack of restrictions is not only related to the fact that many parents believe that teens are old enough to set their own limits in this area. Parents of today's teens are largely "baby boomers" who, having been raised in homes in which visual media served as an important source of conflict and cultural change, express a great deal of ambivalence in their conversations about media. This situation and its implications will be discussed more fully in chapter six.

Why religion?

As noted at the beginning of this chapter, the religious beliefs and practices of media audiences are commonly overlooked in studies of meaning-making practices. This is largely because in sociological and media studies research, religion has been viewed as a residual category: it is widely accepted that religion, while once holding an important and prominent role in public life, has now been reduced to the margins of society due to the increase in rational thought and in scientific knowledge that has been central to intellectual thought since the Enlightenment. Berger (1969) has provided the definitive description of "secularization," as this process has become known:

> By secularization, we mean the process by which sectors of society and culture are removed from the domination of religious institutions and symbols...as there is a secularization of society and culture, so is there a secularization of consciousness. Put simply, this means that the

> modern west has produced an increasing number of individuals who look upon the world and their own lives without the benefit of religious interpretation (pp. 107-108).

Thus not only religious organizations but religious authority has been disappearing, an idea supported through the statistics demonstrating the decline in church and synagogue attendance over the past few decades.

An added problem in cultural studies is the tendency for these works to concentrate on the "popular" as opposed to the "elite" aspects of culture. With its Marxist roots, it is not surprising, therefore, that cultural studies tends to group religion in the latter, "elite" category, and therefore of less theoretical interest.[6]

Yet many engaged in studies of American religion note that religion is not disappearing but is being redefined in its everyday "popular" practices by ordinary people, as is implied in the fact that, despite declines in organized religion, a majority of Americans still identify themselves as religious or spiritual (Roof, 1993). Rather than looking to organized religion, these scholars suggest an exploration of the ways in which religion emerges in everyday life, thus foregrounding the private sphere and the role of religion in self-identity. Authority over religious symbols is no longer thought to be found in religious institutions, therefore, but rests with individuals themselves (Hammond, 1992). Even individuals who still identify with religious institutions approach their decisions about what to believe as a "rational choice" made in a "religious marketplace;" others leave institutions but still identify themselves with their former, as well as a blend of perhaps other, religious affiliations (Iannaccone, 1991; Warner, 1993; Hoge, Johnson, & Luidens, 1994).

[6] See, e.g., the dismissal of religion as "elite" and therefore of less interest in Mukerji and Schudson,1991.

An obvious issue that emerges with respect to the current situation is this: if people spend more time with the media than in formal religious institutions, yet still call themselves "religious," "spiritual," or at least "believing in God," it is possible that their (our) ideas of what it *means* to be religious/spiritual/a believer are drawn, in part at least, from mass mediated sources. These mediated learnings may be part of a rather intentional search, as Eller (1993) has argued in her analysis of the feminist spirituality movement in the U.S. and the importance of popular texts such as Starhawk's *The Spiral Dance* or Marion Zimmer Bradley's *The Mists of Avalon* within it. Yet if we believe that the media, as noted earlier, shape "reality" to a significant degree by establishing and maintaining the frameworks by which it is represented and understood, this also suggests that the role of religion in U.S. culture is both reflected and constructed in relation to these symbolic depictions and hence influences the individual's conception of religion, whether audience members are conscious of this process or not. As Hoover (1997) argues,

> What this new perspective on religion allows for is the construction of religious culture out of a universe of symbols - an inventory if you will - that may or may not include the symbolic claims of the historic religions, and that nonetheless result in legitimate, substantive constructions (p. 286).

Consistent with Hoover, I argue that persons across the continuum of religious ideological positions and practices draw upon this "symbolic inventory" to a significant degree. This is also consistent then with "popular" practices of appropriative religiosity that historians now claim have roots well into the past - even if they did not become an area of interest and study until recently. Certainly it is not difficult to imagine that those Anglo-Americans interested in Native American spiritualities

or ancient Celtic religions rely upon some form of mediated communication (rather than, say, traditional religious socialization through families or religious organizations) to learn about these traditions and to learn of where to meet others who share their interests. And while religious conservatives often attempt to draw a sharp distinction between their own beliefs and those of the "secular humanist" culture expressed in the media, one need only consider the exponential growth of the PromiseKeepers movement and its attendant positive press coverage, the views of which are legitimated each Monday night when famous football players form a prayer circle.

Still, religion might seem an unlikely category for the generation whose members are identified with authors such as Douglas Coupland, who gave the generation its moniker and later penned a self-revelatory book titled *Life After God*. As children of the "baby boom" generation - that group of adults who experienced first-hand the major social upheavals and resultant challenges to institutions that were brought about in the sixties and seventies - today's teens have much less experience in the official institutions of religion. Roof (1993) reports, for example, that nine out of ten baby boomers attended church regularly when they were children, although nearly three-quarters had either dropped out or were attending religious services infrequently by the time they were in their early twenties. While dropping out of church during this stage in life is a common pattern in the U.S., the number of baby boomers who returned to religious organizations after some time is much smaller than in preceding generations. Thus, their children were more likely to grow up in homes in which attending the religious services of churches, synagogues or

mosques was more infrequent than in any previous time - even if they maintained an affiliation with the original religious group.

Most teens feel that their ideas and their parents' coincide in the areas of what values are important in life, including religion, politics, gender roles, racial issues, and the personal aspirations of the teens (Bachman, Johnston, & O'Malley, 1993, pp. 183-185, cited in Chadwick & Heaton, pp. 82-84). Perhaps it is not surprising, therefore, that religion turns out to be an area of curiosity for teens. Most teens report that religion comes up as a topic of discussion with their parents about once a month, while more than 42 percent of teens report that they discuss religion or values with their parent(s) at least once a week and only 16 percent report that they never discuss these topics (Moore, 1992, cited in Chadwick & Heaton, p. 79). Religion, however, is one of the top three topics that teens report they would like to discuss more frequently with their parents, following school work and family issues (Moore, 1992, cited in Chadwick and Heaton, p. 80). Thus as even I was surprised to learn, teens are interested in discussing religion with adults, even if they rarely attend the services of organized religion. Again, this demonstrates the importance of raising the intriguing questions of *popular,* rather than *institutionally-based,* religion and religious identity among teens.

Previous research into the relationship of religion and media has demonstrated the importance of considering religious identifications and affiliations when exploring the meanings television programs hold for their *adult* audiences. In his study of 700 club viewers, for example, Hoover (1988) documented differences in interpretations of the program between evangelicals and fundamentalists. Similarly, Linderman (1995)

discovered differing interpretation strategies evangelicals, Catholics and those not religiously affiliated brought to the viewing of an evangelical film production. Further, Hoover and his colleagues (1995) found that individuals who claim an association with a religious affiliation receive most of their information regarding issues - even issues of religion - from daily newspapers and television. These studies demonstrate that different religious group memberships might predict differences in viewing/interpreting strategies. The current project builds upon and extends these findings to explore meaning-making practices among young persons who would not necessarily be interested in explicitly "religious" (or specifically, evangelical) programming, yet have watched programs available on network television that do in fact communicate a fairly explicit (popular) religiosity.[7] These issues will be discussed in chapters four through seven.

Context for this study

The preceding sections of this chapter have provided a theoretical and methodological grounding for the question of the role of the media in religious identity-construction. In this last section, I would like to set the context further by exploring some of the current information on the social, political, and economic trends in which teens in the U.S. are embedded.

There are just over 50,000 families in the U.S. today, and census data indicate several noteworthy demographic characteristics of the households in which today's teens reside (U.S. Bureau of the Census. 1993, p. 61, cited in Chadwick & Heaton, p. 16). As many studies of birth rates have noted,

[7] A *60 Minutes* (1997) report referred to *Touched by an Angel* as "the most overtly religious dramatic series in television history."

the majority of adolescents in the U.S. are Anglo or caucasian, although this group represents a shrinking percentage of the total number of adolescents as the number of Hispanic, African American and Asian children continues to increase (Chadwick & Heaton, 1996, p. 1). Multiculturalism, therefore, is not only a political (or politicized) issue within educational curricula among today's young people. Their experiences of persons whose racial, ethnic, and consequently religious backgrounds differ from their own are likely to be quite a bit broader than those of their parents. Moreover, more of them are biracial or even multiracial themselves, making "adult" concerns about "the Other" - too often monolithically defined in categories of black/white - seem dated and irrelevant. Issues of multiculturalism and tolerance, as noted earlier, are therefore important large-scale forces that background the current investigation and shape how identity and identifications are discussed.

In addition to racial/ethnic diversity, teens also experience a diversity of familial situations among their friends, as family structures have changed dramatically over the past decades. In 1940, 70 percent of adolescents lived with two parents who had each been married only once. By 1988, however, that figure had dropped to 50 percent. Moreover, the number of children born to a single parent had increased in all racial/ethnic groups, and the percentage of single-parent homes has continued to increase among African American and Hispanic families (U.S. Bureau of the Census. 1993, p. 14, cited in Chadwick & Heaton, 1996, p. 5). The increase in the number of single parents combined with other economic trends, such as the widening gap between rich and poor and declining wages for entry-level positions, has resulted in economic

hardship for an increasing number of U.S. adolescents. While trends such as the increase in parental educational levels, rising rates of female participants in the labor force, and the rising age of marriage have been economically beneficial to some adolescents, the percentage of adolescents living in families with middle class incomes declined throughout the 1980s. Unfortunately, when compared with other industrialized nations, U.S. adolescents experience a higher rate of poverty and a lower rate of assistance from social services (Chadwick & Heaton, 1996, pp. 151-152). Considering these trends, it is not surprising that the number of so-called "traditional" families in which the father works full time and the mother is not in the labor force dropped to under 18 percent by 1990 (Chadwick & Heaton, p. 2).

Moreover, more teens than ever report dealing with depression and related mental problems. More than one-third of all teens had considered committing suicide, with 15 percent of those coming very close to an attempt, and 6 percent actually making an effort. Of those teens who had attempted suicide or seriously considered it, nearly half credited problems at home as a key factor (The George H. Gallup International Institute, 1991, p. 75). These issues suggest that the current dissertation of necessity must consider the changing nature and definition of "the family" as it is experienced by young people today.

The problems of adolescence do not end with economics or destabilized family structures, however. More than 4 percent of today's adolescents have been involved in reported instances of child abuse, with neglect the most common form, followed by physical and sexual abuse (U.S. Department of Health and Human Services, National Center on

Child Abuse and Neglect, 1994, p. 9, cited in Chadwick & Heaton, 1996, p. 53). Moreover, nearly one-quarter of twelfth-grade girls report that they have been sexually abused in their lives (Benson, 1993, p. 28, cited in Chadwick & Heaton, p. 59). Fear of sexually-transmitted diseases, and AIDS in particular, has also increased (The Alan Guttmacher Institute, 1994, p. 41, cited in Chadwick & Heaton, 1996, p. 73). In fact, 23 percent of teens reported that they had changed their behaviors as a result of the AIDS epidemic (Bezilla, 1993, p. 52). Sex, as evidenced in popular culture as well as in other less mediated forms of teen communication such as interpersonal conversation, graffiti and teen chat rooms, is a topic of much conversation among young people. They have never experienced the "sexual revolution" of their parents' generation and do not necessarily associate sex with freedom in the way their parents did as young persons. They are more likely to think of the possible consequences of sex, ironically putting them more in concert with their *parents' parents* generations, although largely for pragmatic rather than moral reasons.

As I will discuss the survey results of teen audience behavior in a subsequent chapter, I will not go into detail here. However, one interesting finding that seems particularly relevant is that while only 13 percent of U.S. teens believe that religion has a great influence on teens, in contrast, 41 and 32 percent of teens respectively believe that music and television have a great influence on teens. Does this demonstrate the "power" of the media and the lack thereof for religious institutions? Not necessarily, although it does demonstrate the seemingly contradictory phenomenon which finds both teens and adults assuming that the negative influences of the media are of much greater consequence to

others than to themselves.

When teens were asked the less hypothetical question of who was the most influential person in *their own* lives, for instance, the largest group chose their mothers, followed closely by fathers and then both parents. This supports the current study's emphasis on the family context as a primary venue for the development of religious identity. Moreover, only 12 percent of teens reported that their friends were the most influential persons in their lives (The George H. Gallup International Institute, 1991). This problematizes the self-reporting of believed influence while also providing evidence of the public discursive approaches to the media, both of which figure importantly in a study such as the current project.

A large Canadian research project on teens concluded that many teens have few adults in their lives who model meaningful and fulfilling lives. As a result, they suffer a crisis of belief that encompasses belief in themselves, in their relationships, in God, and in the future (Lefebvre, 1997). I will argue in this study that this crisis has resulted from a combination of the factors outlined above, in addition to the trends of religion in American life which point toward a continued questioning of religious institutions.

It practically goes without saying that teens today experience a proliferation of entertainment options afforded through the mass media, which might be broadly defined as including not only television, film, radio and newspapers but also video games, music on CDs, and the Internet. In the past decade as media ownership regulations have been relaxed, an unprecedented series of mergers and acquisitions continually

altered the map of U.S. media ownership. While it is clear that the proliferations of channels available on television resist analysis as a totalizing hegemonic force within society, it is also clear that hegemony continues to work in the media as certain issues are highlighted while others ignored altogether. One issue, for instance, is the paradox of increased globalization and yet the argubly simultaneous narrowing of perspectives on world issues. CNN, as it proudly claims, is now the most-watched news network around the world. Yet most of the correspondents are either North Americans themselves or educated in the U.S., fitting their work easily into the North American conception of what is newsworthy.

Another interesting paradox of the media is that while the multinational and profit-driven media are seen as remote and perhaps "dangerous," their interest in drawing audiences has actually resulted in an unprecedented increase in the research into what audiences want. In media studies, of course, there is great debate about whether or not what the audience *wants* is what it in actuality *needs,* but no one can deny that the media attempt to reflect their audience's values - whether those be deemed "positive" or "negative" - to a greater extent than ever before in the search for increased market share.

While their parents grew up with the introduction of television into their neighborhoods and homes, today's teens take for granted the niche cable channels of the Black Entertainment Network, MTV, Nickelodeon, and even the Science Fiction channel. Thus while it is still true that programming choices are limited, as critics have often complained, it is certainly also more possible now than ever before for

those critics (and their teenage children) to construct their own programming schedule of uplifting, wholesome programming, however they define that.

An Outline of Subsequent Chapters

In the next chapter, I go into further detail about the origins of the notions of the "self" and the relation of the concept to Western ideology, modernity, and postmodernity. I also critique the ways in which questions of identity have been addressed in the various literatures of media studies. Finally, I make an argument for the connections between the discourses of "meaning" and of religion in the U.S., illustrating this process through an analysis of the practices of meaning-making observed in a case study of a 16-year-old girl and her single-parent father.

In chapter three, I describe more fully my methodological approach and its theoretical underpinnings. I also reflect there upon the practical issues that emerged in conducting the fieldwork, each of which influenced the shape of the final product and my own understandings of the knowledge produced in the process. The young people who served as "case studies," as well as other teens and parents of teens, will be introduced in this chapter, as well.

Chapter four delves more fully into the discussion of adolescents as a media audience (or audiences), reviewing the current literature in the field which deals with both adolescents in peer groups and in families. Teen discussions on both parents and peers illustrate the importance of incorporating both contexts into any meaningful analysis of young people and their identity-construction projects.

Chapter five extends the work of the previous chapter, offering a case study of one teen. A detailed description of her identity-construction, with particular attention to its implications for her understandings of religion, provide a sense of what it might mean when a teen calls herself "religious" today.

In chapter six, I reflect back on the concerns that parents express with regard to their teens' media use and their development of meaningful belief systems, whether religious or otherwise described. I note the differing approaches to both religion and media, and how these two intersect for parents. I relate how they discuss the various tactics they employ to manage the role of the media in their teens' lives and their ambivalence about this. I also analyze the differences in the cultural environment between today and the time when today's parents of teens were teens themselves.

Chapter seven further explores what I believe it means when teens say they are "religious" through an analysis of the relationship of their conversations with the public discourses of religion which emerge in relation to the popular television program, *Touched by an Angel.* I discuss the common themes in these discussions, while also critiquing the limits of tolerance as it is expressed in terms of religious difference among young people.

In chapter eight, I conclude with a discussion of the implications suggested by the differences in interpretation strategies I observed between teens and their parents. These include changes in the cultural status of religious symbols and institutions and in patterns of reception among teens. I also describe the rise of self-reflexivity observed among teens and

its relation to an increasingly therapeutic culture, concluding with implications for future research.

Chapter Two

The 'Self' in Meaning-Making: A Critical/Cultural Approach to Identity-Construction at the Site of the Audience

The past few years have seen the development, refinement and differentiation of approaches to research on the "active audience" in media studies. The researcher's approach to the processes of gathering data and its interpretation, the relationship of the researcher to those interviewed, and even a project's aims differ across paradigms to such an extent that it is difficult to conceive of the current body of research as one that is "cohesive" or sharing in common concerns, as Guba and Lincoln (1994) and Ang (1989) have pointed out. Still, it might seem that the current widespread interest in identity as a rubric for investigation may offer bridges for connection. This chapter, therefore, sets out to review the traditions and current debates in audience research around issues of identity for points of commonality and distinction. I will discuss how the critical/cultural studies approach has guided my own understanding of the identity-construction process and how my fieldwork, in turn, has challenged certain assumptions of the categories through which identity might be explored. I begin with a discussion of identity and the concepts of the self which shape this dissertation. Next, I review the paradigmatic approaches to questions of media and identity, concluding with an extended discussion of the differences I see between the critical/cultural studies approach and those that have preceded it. I conclude with an analysis of a case study that emerged in my fieldwork and what I believe

are its implications for continued explorations into the role of media in religious identity-construction.

The Concepts of Identity and the Self

What are the roots of the current understandings of identity and the self that shape U.S. media research? Bellah and his colleagues (1985) have argued that the contemporary concern for the self may be traced to the earliest days of America, as biblical and republican traditions met in the form of Puritans and early colonists who struggled for freedom from governmental restraint. Habermas' (1992) analysis suggests a much earlier timeframe, noting that the Greek *polis* was based on the autonomy of the master of the household who was freed from the mundane tasks of everyday life to participate in governance. Lash and Friedman (1992) argue for what might be considered the "received" tradition, dating such concepts to the rise of modernity and Enlightenment thinkers. With the caveat that each of these analyses privilege a European tradition of scholarship and history, I argue that it is important to consider this history for its influence in shaping the contemporary struggles over the definitions of identity as they have emerged in media research.

All of these scholars share the conviction that our current notion of "self" was shaped by the widespread cultural changes of the 15th century. Prior to this time in Europe at least, identity was conceived of primarily in relation to both a deity and a king and thus it is surmised that there was little interest in or attention to an autonomous self. Yet as the authority of the Catholic church was undermined in the Protestant Reformation through both the spread of scientific knowledge and the increasing literacy

of the common people, philosophers began to question both religious and ultimately governmental authority over the individual. Foreshadowing Enlightenment thought and in particular Feuerbach and Marx, Puritan religious dissenters such as the Quakers, the Diggers, and the Levelers of the 15th century insisted upon a form of social justice, arguing that religion as it had been practiced had diverted persons from asserting their political rights in relation to the monarchy. Locke's argument that a leader's legitimacy rested upon a "social contract" that demanded his protection of the interests of his subjects was then influential in the English Revolution of 1688.

Enlightenment thinkers of the 18th century, led by Kant, argued for a subject capable of making cognitive, moral, and aesthetic judgments autonomously. The ability to make these choices was what made one human, and by extension, an individual. Bringing the empirical approach of science into philosophical argument, Kant and later Hume argued for the importance of the *subject* 's perceptions of the world as the mediator for the development of reason, thus foregrounding a subjective position and encouraging a concern with "self-mastery and self-domination," as Lash and Friedman (1992) note. Geertz (1979) describes the resulting concept of the individuated self in this way:

> The Western conception of the person as a bounded, unique, more or less integrated motivational and cognitive universe, a dynamic center of awareness, emotion, judgment and action, organized into a distinctive whole and set contrastively against other such wholes and against a social and natural background is, however incorrigible it may seem to us, a rather peculiar idea within the context of the world's cultures (cited in Sampson, 1989, p. 1).

The self is thus the *center,* or the site of "primary reality," as Sampson (1989) points out, a being with the powers of awareness, emotion,

judgment and action.

The notion of a "centered" self with both the ability and the responsibility to make judgments well and to the benefit of both self and others is central to the Liberal political tradition of Milton, Locke, Bentham, Jefferson, and later, Mill, all of whom advocated a goverment that would not restrain individual reason. This political philsophy undergirds and legitimizes the rise of capitalism as the socioeconomic order that replaces feudalism. It is worth noting that the pragmatic application of this tradition in the U.S., reserving some ideas from English governance, assumed that participation in governance was limited to property owners. Many critiques of the early U.S. political system's claims to be "for all" have been articulated already and need not be mentioned here (See, e.g., hooks, 1990; Gates, 1992). However, it is important to note that the idea of the "self," as it has been employed in political philosophy, refers to a naturalized subject position of the property owner, a position at that point in history restricted by class, race, and gender. Thus the ideas of the "self" and of the "individual" are not ideologically neutral, but are inherently related to the Western cultural context in which they developed. Recent approaches to notions of the self and the broader questions of collective identity, as I will discuss later, have been conscious and critical of this ideological construct.

Writers of the "high Modernist" period such as Marx, Weber and these political philosophers were concerned with a paradox of the Enlightenment: how could it be that the Industrial Revolution, as it provided freedom from the unpredictabilities and oppressions of agrarian life, could also submit people to new forces of domination and control?[1]

[1] Lemert (1993) poses this question in his introduction to the social theory of Modernism.

While on the one hand the political revolutions of the time seemed to provide increased freedoms for individuals, on the other hand, people did not seem capable of escaping the oppressive conditions of their lives. To explain these seeming contractions, Weber looked to the realm of ideas, exploring the links between history, individual motivations and societal material experiences. Marx looked to the realm of the material relations between economic classes. Yet perhaps the scholar with the most important influence in the development of the theory of the "centered" self was Freud.

In contrast to the rationalism inherent in theories of Weber and Marx, Freud addresses the *irrational* motivations which are largely inaccessible to individuals yet which govern their behaviors. This inaccessible part of the personality is the Id, which strives to fulfill the "Pleasure Principle" (e.g., seek pleasure/avoid pain). The Superego is that which keeps the Id in check, and is the internalized voice of societal norms learned, Freud argues, from parents. The Ego, Freud writes, "is the sense-organ of the entire apparatus...receptive not only to excitations from outside but also to those arising from the interior of the mind (cited in Jacobs, 1992, p. 58)." The Ego is that which mediates between the external world, the Id and the Superego, and is responsible for achieving the "Reality Principle," synthesizing and unifying the individual's mental processes. It is important to note that Freudian psychoanalysis is based on the assumption that the irrational and inaccessible in one's unconscious are fundamentally related to one's behaviors. These ideas are repressed by the Ego, but return in the form of dreams and neuroses which guide behavior, although the conscious Ego is not aware of this fact.

Freud's concept of the Ego understandably held great appeal in the early United States, for it seemed to affirm the notion of the autonomous and self-governing individual who is ultimately responsible for his or her own actions. Tocqueville (1969) describes this quintessential American outlook on the world as "individualism," a term he popularized. He writes:

> Individualism is a calm and considered feeling which disposes each citizen to isolate himself from the mass of his fellows and withdraw into the circle of family and friends; with this little society formed to his taste, he gladly leaves the greater society to look after itself...there are more and more people who, though neither rich nor powerful enough to have much hold over others, have gained or kept enough wealth and enough understanding to look after their own needs. Such folk owe no man anything and hardly expect anything from anybody. They form the habit of thinking of themselves in isolation and imagine that their whole destiny is in their hands (p. 506).

We can see that this analysis still holds a great deal of relevance today, and in part explains the fascination with the interior workings of the self that have been the bulk of psychological research in this century. Yet the concept of the "centered self," so important to the project of psychology, has increasingly come under scrutiny (See, e.g., Shotter & Gergen, 1989). In part, this has happened as a result of the challenges to the "centered self" introduced in critical theory and in poststructuralism.

Critical theory and the "self"

The concept of the individual as a source of authority, able to think and reason and use these abilities to bring about a better world, was not only foundational for psychology, liberal democracy and capitalism; it undergirds Marx's thought to a significant degree, as well. Marx's notion of a non-oppressive society rests upon the belief that workers may use their

autonomous reason to become conscious of their exploitation in labor relations and revolt against it, thereby bringing about a society of equality and fairness; "an autonomous ego and an emancipated society reciprocally require one another," as Habermas (1979, p. 71) states. Yet consciousness is always limited by one's experiences and one's location in historical context, as Marx and those embracing his school of thought acknowledge. One's "self," in this model, is therefore determined, or limited, by one's relationship to the modes of production, which in turn have been determined by a long history of the relationship between the owners of production and workers.

Marx was of course concerned with the exploitive relations in the industrial situation and recognized the historical origins of the patterns he observed in the industrial revolution. In this approach, history is understood as encoded into a culture's symbolism, or "superstructure," and provides a framework within which the individual perceives reality. As C. Wright Mills (1963) writes:

> The first rule for understanding the human condition is that men live in a second-hand world. The consciousness of men does not determine their existence; nor does their existence determine their consciousness. Between the human consciousness and material existence stand communications and designs, patterns and values which influence decisively such consciousness as they have (p. 375).

Marxist thought, therefore, challenges the notion of the purely autonomous self, requiring an understanding of historical and cultural context in the analyses of the relations of the individual and society. As Sampson (1989) points out regarding those working within the Marxist and neo-Marxist traditions:

> Critical theorists argue that there is an essential interpenetration of society and the individual that warrants our approaching with

> scepticism [sic.] any view that makes the individual a transcendent entity. We do not begin with two independent entities, individual and society, that are otherwise formed and defined apart from one another and that interact as though each were external to the other. Rather, society constitutes and inhabits the very core of whatever passes for personhood: each is interpenetrated by its other (pp. 3-4).

Thus critical theory participates in a "decentering" of the self as neither the individual nor the society have primacy. Sampson (1989), paraphrasing Giddens' structuration theory, notes:

> The person is the mediated product of society and also, in acting, reproduces or potentially transforms that society. People can transform themselves by transforming the structures by which they are formed (p. 6).

It is important to note, therefore, that the agency of the "self" is not completely lost in critical theory, but that a *dialectic* between the individual and society is emphasized to the extent that the autonomous, "centered" self of psychology is challenged as an ideological concept that maintains current relations of power within society.

As the concept of the agentive individual is not completely lost, the idea of an individual's *consciousness* is also not inconsistent with critical theory or with the "decentered" self in this model. Gramsci is credited with re-introducing human subjectivity into Marxism, bringing to the fore certain readings of Marx (particularly the earlier, Hegelian-inspired writings) which discuss the role of ideas and consciousness in both Marx's proposed revolution and in the forces which mitigate against revolutionary consciousness. Gramsci argues that the rule of one class over another cannot be explained by material power alone, but is established and maintained through 'hegemony,' the process by which the norms and values of the ruling class are naturalized and conventionalized

as the norms and values of all. Gramsci's notion of hegemony builds upon Marx, as can be illustrated in this famous passage from *The German Ideology* (1974):

> The ideas of the ruling class are in every epoch the ruling ideas, e.g., the class which is the ruling material force in society, is at the same time its ruling intellectual force... Insofar, therefore, as they rule as a class and determine the extent and compass of an epoch...(they) rule also as thinkers, as producers of ideas, and regulate the production and distribution of ideas of their age: thus their ideas are the ruling ideas of the epoch... (this class therefore is) compelled...to represent its interest as the common interest of all the members of society...to give its ideas the form of universality, and represent them as the only rational, universally valid ones (pp. 64-66).

Yet hegemony, as it has come to be understood within media studies, is not limited to the realm of ideas. Raymond Williams (1977) defines the term this way:

> 'Hegemony' goes beyond 'culture,' as previously defined, in its insistence on relating the 'whole social process' to specific distributions of power and influence. To say that 'men' define and shape their whole lives is true only in abstraction. In any actual society there are specific inequalities in means and therefore in capacity to realize this process...It is in recognition of the wholeness of the process that the concept of 'hegemony' goes beyond 'ideology.' What is decisive is not only the conscious system of ideas and beliefs, but the whole lived social process as practically organized by specific and dominant meanings and values (pp. 108-109).

While it frames the individual's perception of reality and obscures alternative approaches, hegemony is never fully secure, and thus ideological struggle ensues when the view that once was universally accepted undergoes widespread challenge. Therefore a revolution would require not just force but a change in ruling ideas. Gramsci argues that opportunities for hegemonic change come about during periods of social and political upheaval, such as that which occurred after the first World

War. During such times the 'normal' exercise of hegemony, in the hands of the dominant forces of society, is sufficiently undermined to allow the possibility of a new basis for consensus.

Gramsci's (1971) value to the current project lies in his simultaneous foregrounding of material conditions and structures of domination in culture while suggesting the importance of the symbolic, 'superstructural' realm in shaping how people understand their lived experiences, through ideology. The definition of culture that emerges therefore shares common ground with the configurational approach of Ruth Benedict (1934) and others in cultural anthropology. As Ortner (1984) writes, "The problem is not that of the system telling lies about some extrasystemic 'reality,' but of why the system as a whole has a certain configuration, and of why and how it excludes alternative possible configurations." Culture is therefore understood as constraining and yet open to reshaping through the actions - whether intentional or not - of human agents.

Poststructuralism and the "Decentered Self"

While I have argued that it is possible to employ the concept of the "decentered self" in a general way within critical theory, the term itself is most frequently associated with poststructuralism's critique of the subject, notably in the writings of Derrida (1978). Much as structuralism foregrounded the centrality of language, poststructuralism posits that the subject or individual self is not the *source* of meaning but is a *product* of signification. The self, in this model, is constructed in language. Yet unlike critical approaches, poststructuralism rejects a notion of individual

consciousness while simultaneously reserving the notion, borrowed from Freudian psychoanalytic approaches, that the self's identity is not immediately accessible to us.

Employing Saussure's (1917) idea that signs and that which they represent can only be understood in relation to what they are *not,* poststructuralists view the person as an outcome of a complex and ongoing process of distinguishing the self from what he or she is not. Yet simultaneously, the person also appears in the immediacy of the present to be "whole" and unique. This is the essence of Derrida's concept of *differance,* which combines Saussure's notion of difference with Freud's concept of deferral (the paradox of a self appearing cohesive while resulting from ongoing processes over time). This process is then related to the *decentering* of the self, as the symbolic system through which subjects are fixed in place is seen as central (rather than the individual subject). The subject, then, is fluid, always changing in its presentation, and yet still fixed in place by (and therefore understood by others within the context of) the culturally inscribed symbolic order.

While I find the concept of the self as constructed in relation to the symbolic sphere to be helpful as it challenges the notion of the Ego striving toward a unified, "integrated" self, there is an important aspect of poststructuralism that I reject. I am not sure that the integrative Ego and the conscious self need be equated and then dismissed. Critical/cultural approaches are able to bridge this problem not through the complete rejection of the concept of identity, as Derrida proposes, but through a reconceptualization of identity in relation to ideology.

Because I am interested in employing critical theory in the present work, I

see it as important to maintain the concept of the conscious individual, agentive yet constructed in language and hence whose agency is always limited by the symbolic and material context in which he or she is located. Derrida can be useful in this theorization, as Sampson (1989) explains:

> Derrida's thesis is that presence is always already mediated by the absent trace; thus, self-consciousness is not a direct and unmediated experience but rather is an indirect and always already mediated experience. This way of understanding personhood and consciousness permits a key role for social and historical traces to enter and structure the very experience of consciousness and of self, even as those traces are unavailable to presence and awareness. As Coward and Ellis (1977) observe, this perspective leads us to see ideology - specific sociohistorical traces contained within the language system - permeating the very core of personhood. Ideology is not a garment that one puts on and removes at will: ideology constitutes the person as a subject in the first place (p. 13).

Consciousness, therefore, is preceded by the process that constructs the person within the symbolic system, Sampson argues.

I think it is helpful to consider the conscious individual as able to actively engage in the production of meaning-making, albeit within certain restraints of cultural understandings that are shaped by discourse. I use the concept of *discourse* with reference to the culturally specific frameworks within which all narratives are constructed, whether they be mass mediated or emerging in interpersonal communication, which bring certain elements of the hegemony into an acceptable and plausible framework for members of the culture. Discourse, as O'Sullivan and his colleagues (1983) write, is "the social process of making and reproducing sense(s) (p. 73)." Thus, while the language in which a discourse operates may be said to be infinite, a discourse is limited by the specific context of the structure of social relations in which it occurs. Further, O'Sullivan and colleagues note:

> we establish and experience our own individuality by 'inhabiting' numbers of such discursive subjectivities (some of which confirm each other; others however coexist far from peacefully). The theory of discourse proposes that individuality itself is the site, as it were, on which socially produced and historically established discourses are reproduced and regulated (p. 74).

Therefore, the subjective experience of the individual is the location at which conflict between different ideological positions occurs and is resolved, if only temporarily, through the ways in which individuals participate in discourse. This is in fact where use of the concept of identity becomes relevant for Hall (1996), as noted in the introduction:

> I use 'identity' to refer to the meeting point, the point of *suture*, between on the one hand the discourses and practices which attempt to 'interpellate,' speak to us or hail us into place as the social subjects of particular discourses, and on the other hand, the processes which produce subjectivities, which construct us as subjects which can be 'spoken.' Identities are thus points of temporary attachment to the subject positions which discursive practices construct for us...The notion that an effective suturing of the subject to a subject-position requires, not only that the subject is 'hailed,' but that the subject invests in the position, means that suturing has to be thought of as an *articulation*, rather than a one-sided process, and that in turn places *identification*, if not identities, firmly on the theoretical agenda (pp. 5-6).

Thus we retain the conscious self who is able to construct an identity within the limits of discourse as he or she *articulates* a particular discursive construction.

As noted earlier in this chapter, however, this approach to the self and the question of identity has been virtually nonexistent in the empirical work of media audience studies. Thus I would now like to explore how the intellectual traditions of Freudian psychoanalysis and symbolic interactionism have shaped the ways in which most questions of

media and identity have been framed in U.S. research.

Analytic Ego and Cognitive Developmental Psychology: Identity as the Developing , Autonomous Ego

While the notion of the "essential," "centered" self and the field of psychology itself has come under scrutiny in the past years, media studies of identity in the U.S. still often assume a stable and developing subject for analysis (For a similar critique, see e.g., Pfister & Schnog, 1997). The works of Jean Piaget and Erik Erickson in the middle of this century have been particularly influential in the identity literature of mass communication research, for example. Piaget (1969) argues that there are developmental stages that are reached chronologically in the lives of individuals. While these stages are generally related to the child's age, they are primarily reached as individuals achieve certain skills. Erickson (1968), also basing his argument on stagewise development, theorizes that the ego identity emerges in the fifth, adolescent, stage of psychosocial development. Erickson claims that each person undergoes an "identity crisis" during this stage of life which forces people to make irreversible decisions regarding self-definition. As many of his critics assert, Erickson's viewpoint assumes that it is up to the *individual* to choose between the options society has to offer, thus placing the responsibility for making the "best" choices with the individual.[2]

Psychological and stagewise developmental explanations of identity have held a great deal of sway throughout U.S. culture, and thus it is not surprising that psychological frameworks have greatly influenced the bulk of research into young people's identity in media studies, as I will discuss

[2] For an excellent critique of Erickson, see Slugorski & Ginsburg, 1989.

in chapter three. The implicit assumption behind this research is made explicit by Strasburger (1995), who argues for the urgency of his inquiry into television's deleterious effects on adolescents. As he notes:

> young teenagers may be the most susceptible population to television's hidden themes and messages because their identities and attitudes are evolving and more malleable (p. 12).

The primary problem with this approach, I believe, is the extent to which it privileges media content in constructions of the self to the exclusion of other environmental factors. For example, Strasburger (1995) argues at one point:

> Because television contains a fairly uniform set of social messages, television viewers' social schema tend to be whittled down to a small set of similar beliefs. For example, because the number of doctors, lawyers, and policemen are overrepresented on American television, adolescent males may develop the mistaken impression that only professionals have value in adult society, not blue collar workers. Likewise, adolescent females may be selling themselves short on the basis of what they see on TV(pp. 10-11).

This analysis completely overlooks the ways in which the capitalist system structures such valuations of professionals over blue collar workers and how the maintenance of both class and gender inequalities serves the interests of those with greater economic and social power. It also overlooks, then, the ways in which these structural positions influence one's sense of identity and also limits one's options with regard to which identities are possible or may be envisioned by the subject. Further, by assuming the individual is the unit of analysis, the approach fails to explore the relationship between the symbolic forms circulated in the media and the *structures* of society. This analysis, as is true with many in this approach, is flawed by its inattention to cultural context, assuming a

reductionistic direct relationship between media content and the formation of the sense of self.

In addition to these studies which have been termed 'direct effects,' influences of analytical ego psychology can also be found in the work of those researchers who have argued that the media's effects on the individual are not strong but rather quite limited. These researchers focus on the uses of media and the gratifications this use offers to its consumers. Scholarship in this approach has asked: "what does the behavior of young people regarding media tell us about their knowledge, beliefs, and values?" Rosengren and Windahl (1989), for example, posit that children have different media needs and preferences based upon the developmental stage which they are experiencing at the time. Inserting media practices directly into Piagetian theory, they argue:

> During the course of cognitive development children thus become less dependent on their immediate perception, and their ability to deal with multiple dimensions of phenomena expands...We learn from cognitive stage theories how children obtain more and more means and resources for utilizing the media (p. 57).

While I appreciate the way in which this approach complicates the relationship of media and audiences, particularly as it looks not just at content but the practices of media consumption, I argue that one problem of this model is that it tends to tie "needs" rather unproblematically to stages of development without considering other reasons why people might be motivated to watch or listen. As we saw with Erickson and others in psychology, this approach assumes that it is the *individual* who is responsible for finding ways to have her or his "needs" met - and thus if he or she fails in this process, it is the individual who is labeled pathological. The ability of society to *allow for* the meeting of needs is a

question not asked in this framework.

I do think that analytic ego and developmental psychology contribute important understandings of the cognitive development adults achieve, and I will address the learnings of these studies more fully in chapter 4. The problem in rather unproblematically adopting these frameworks for media study is that they tend to reinforce rather than challenge an individualistic orientation, thus reducing questions of media and identity to a rather simplistic relationship of media content and its "impact" on individuals. Ironically, while these psychological approaches are rather easily coopted into conservative positions, this approach is also common among those outside of media studies who radically critique the media for its negative stereotypes and their assumed negative impacts on individuals. The fact of this unlikely shared approach to media demonstrates the extent to which psychological approaches have shaped understandings of the relationship of media and identity. Yet there are other ways in which this relationship may be conceived, as we shall see.

Symbolic Interactionism: Identity as Self in Relation to Others

Symbolic interactionists have explored the questions of media and identity primarily through their development of research methods which are based in the assumption of a discrete subject and object, but also through their attempts to theoretically link individuals with larger social processes in the shaping of identity. Their research has addressed how a sense of self arises in relation to others, and how communication plays a role in that process as well as in the process of societal consensus. Cooley's (1902) notion of the "looking-glass self" points to the ways in which

individuals were constructed in relation to how they believe others see them, while Mead (1910) argues that the self can be studied as an object, pointing to a scientific approach to interaction rather than the more mental orientation of Cooley.

One of the limitations of symbolic interactionism is that the theory of *interaction* has generally been more developed than the theory of the role of *symbols* within it. As Denzin (1992) notes, theorists in this paradigm insist on seeing language "as a window into the inner life of the person" rather than as a sign within a larger, unstable and contingent symbolic system as the structuralists and poststructuralists have argued (p. 2). Thus while symbolic interactionism has generally been associated with the development of sociological studies in the U.S. (as opposed to psychology), the symbolic interactionist still views the subject of the self as *centered.* The focus of the research is on the self as a subject position from which to act and speak in relation to others (as opposed to Marxist and poststructuralist concepts of the self spoken by and through culture). This assumption has been influential in terms of the development of and approach to research methods which have foregrounded subjective experience and advocate taking research subjects "at their word," allowing theory to "emerge" from this (See, e.g., Park, 1925; Blumer, 1969; Glaser & Strauss, 1967). Because the theory of society in this paradigm remains truncated, analyses of the relations between the media and identity-construction processes, like the psychological paradigm, remain individualistic.

American Cultural Studies and Uses and Gratifications: Identity as Actively Constructed

To a large extent, the theories of symbolic interactionism put forth at the beginning of this century have been reclaimed in what has become known as the "American cultural studies" tradition of media research. The American cultural studies school is characterized by an embrace of the humanism and liberal-pluralism of Dewey and to a lesser extent, the functionalist assumptions adopted from sociology's research beginning in the middle of this century. Thus such researchers as Carey (1989) and Newcomb (1994), for example, tend to be interested in questions of how communication media assist in providing cultural cohesiveness and stability rather than in questions of conflict or change (See also Dayan & Katz, 1992). Scholars in American cultural studies continue the symbolic interactionist's commitment to ethnographic research, maintaining with it the latter's separation of objective and subjective positions and thus a "realist" approach to ethnography. Thus in an important sense, the concept of identity in this approach is linked to this ability of the research subject to maintain an autonomous subjective position.

The symbolic interactionist model of interpersonal relations is important in the theorizing of the relation between media institutions and media audiences in the American cultural studies approach, as White (1994) notes:

> both media producers and users are negotiating with each other to obtain responses that coincide with the intentions of the respective actors. People creating the media seek to establish identities through favorable responses from their professional peers and from the audience (directly or indirectly)...Audience members may take identity models directly from the media or media may supply the identity achievement strategies and the sources for validation of identities (p.

14).

The assumption behind this approach is that cohesive identities, both individual and collective, are both desirable and achievable, and that the media serve as an integrative agent in social relations. This model is more sensitive to the structural issues of the media/identity relationship than either the 'direct' or 'limited effects' models, as it introduces the notions of genre and professional practices, highlighting institutional prerogatives in what is ultimately represented in media content (See, e.g., Schudson, 1977; Tuchman, 1978). This opens the question of identity beyond the audience level to the identities of media producers and thus helpfully expands the relationship of media and identity beyond individualistic models. Thus research into identity and media in the American cultural studies tradition significantly contributes an analysis of the *motivations* for why people participate in media practices which maintain the status quo at both individual and institutional levels. Too often, however, researchers take these explanations on their face, failing to place them in the larger contexts which might interrogate how those motivations are structured and maintained ideologically. This results in part from the view that "class" is seen as a less relevant category in American analyses, and thus even when Marxist-based theories are introduced, there is often little connection between these theories and the analyses that actually emerge from the empirical data. This is a primary criticism I have of Lull's (1995) work, for example.

Research into identity within the American cultural studies model, based in these assumptions, has focused on how the media, and specifically television, serve as a positive force to maintain cultural

identity or personal identity. Yet another approach central to this tradition has looked at individuals and the role of media in their social relations; for example, how specific social groups use television texts as a resource in conversation to validate their own identities, and how the selection of television programs and the technologies themselves define and confirm masculine and feminine identities and familial roles in the home. As Lull (1988) argues: "Audience members are family members, too, and their identities, interests and roles are articulated, acted upon, and played out in routine activity at home (p. 237)." This statement, with its emphasis on social relations, is in keeping with the symbolic interactionist tradition yet emphasizes the larger question of familial expectations and interests which are linked to the social sphere.

Although what is known in media studies as the 'uses and gratifications' tradition, as noted, grew out of a critique of the 'direct effects' of media, its approach to identity has shared some common ground with the American cultural studies tradition in its concern for cultural cohesiveness and humanism. In both the American cultural studies and 'uses and gratifications' traditions, therefore, the "self" is viewed as "centered," an active agent in the meaning-making process of media reception. *'Uses and gratifications' foregrounds individual agency*, as seen in this foundational statement from Katz (1959):

> ...less attention [should be paid] to what media do to people and more to what people do with the media. Such an approach assumes that even the most potent of mass media content cannot ordinarily influence an individual who has no 'use' for it in the social and psychological context in which he lives. The 'uses' approach assumes that peoples' values, their intersts, their associations, their social roles, are pre-potent and that people selectively 'fashion' what they see and hear to these interests (p. 2).

The 'uses and gratifications' paradigm contributes a sense of the complex processes of what motivates an individual audience member to employ media to either directly or indirectly meet needs, thereby drawing upon Parsons and functionalism in its conceptualization of the focus of research. Research in these traditions is framed as "how individuals make use of the forms and content of media for making decisions or inferences about the nature of their world," a "domain that can be usefully explored through an interpretive framework" (Lindlof & Meyer, 1987). Yet in its interest in asserting the agency of the individual, it differs from the critical/cultural tradition's concern for the dialectic of the individual and society.

The primary difference I see between what I have called the American cultural studies and the uses and gratifications traditions is, first, the former's attention to structural issues rather than the uses and gratification scholar's primary focus on individuals as audience members, and later, the more intentional grounding in the theories of postmodernism and social constructionism among the American cultural studies school - although not always successfully, as we will see.

The notion of social constructionism as articulated by Berger and Luckman has been incorporated into the some of the recent writings of the American cultural studies school, for example, problematizing the "coherent self" assumed in the earlier work of symbolic interactionists (Berger & Luckmann, 1967). To the extent that constructivism and contingency emerge as themes, it would seem that the American cultural studies tradition is responding to the centrality of the metaphor of language and text which emerges in the "linguistic turn" which fuels

structuralism and poststructuralism. One would think that such a turn would prompt a challenge to the stable and autonomous subjective position inherent in the realist ethnography approach of this research tradition, but instead this school adopts the postmodernist concept of "appropriation," thus retaining the agency of the individual actor and continuing the tradition of neglecting the question of from whence the symbols appropriated have come. This notion of appropriation and social construction emerges in psychologist Kenneth Gergen's (1991) argument of how communication technologies create a condition of "social saturation" and increase the "postmodern plight" of a loss of a communal-based "self." He writes:

> Under postmodern conditions, persons exist in a state of continuous construction and reconstruction; it is a world where anything goes that can be negotiated. Each reality of self gives way to reflexive questioning, irony, and ultimately the playful probing of yet another reality. The center fails to hold (p. 7).

Gergen writes of a condition here and throughout his book that is rather specific to the middle class experience in Western cultures, a privileged class positioned in such a way as to allow for such choice-making and negotiation. His lack of reflexivity regarding the cultural context which informs his assumptions is consistent with what we see in media studies which cite his work, such as that of Grodin and Lindlof's (1996) recent influential book. The empirical work in Grodin and Lindlof's edited volume focuses on how people use mediated forms in the construction of self, looking at talk shows, self-help books, magazines, and rap music. Several of the writers in this text have conceived of identity within traditional psychological frameworks: self-esteem, self-fulfillment,

and media as therapy.[3] Most take a 'uses and gratifications' approach, assuming that subjects need certain things in order to achieve an identity, and may meet those needs through media uses. In this way then, American cultural studies merges with 'uses and gratifications' here, with identity placed into a functionalist framework of social cohesion and questions focused on the role of the media in maintaining this cohesion and self-integration through the construction of individual identities.

The Grodin and Lindlof text also highlights the difficulty of integrating the postmodern challenges to notions of the self with a 'uses and gratifications' or 'symbolic interactionist' framework. For while the introduction to the volume briefly discusses the postmodern themes of the "decentred self" and the role of the communication media in the process of the collapse of time and space distinctions, these themes are not addressed in the empirical chapters that follow. It seems to me that this is an indication, in this model, of the importance of the *centered* subject - both for research purposes and in keeping with the general U.S. resistance to the critical/cultural and poststructuralist approach to the symbolic realm as that which "speaks" the subject. There is thus a subtle contradiction inherent in the American cultural studies' embrace of poststructuralism's approach to the "self."

Because of the attention to the social constructedness of social life and of individuals, the researchers in the American cultural studies approach are able to address issues of gender, race and class as aspects of "difference" which play a role in the *choices* one must make. This is seen, for example, in Haraway's (1985) work on "fractured" identities, which

[3] See, e.g., the chapters by Simonds, Priest, Carbaugh, McNamee, and Turkle, in Grodin and Lindlof (1996).

does acknowledge that positioning occurs based on race, gender, or class, yet argues that some differences are "playful" and ironic. Indeed, some feminists, and I would include myself in this, would consider some of the 'feminist' approaches of the American cultural studies school to fall into a tradition of liberal feminism which tends to seek change *within* the social system rather than challenging the system and its mutually beneficial systems of oppression. Thus again, the American cultural studies' interest in individual agency as opposed to its *limits* is what primarily distinguishes it from critical/cultural approaches in relation to questions of identity.

Critical Theory and Critical/Cultural Studies

Perhaps it will strike some readers that the definition of critical theory I offer here differs from that which is often critiqued in the American cultural studies and related traditions. Indeed, some critics of Marxism within media studies have dismissed Marxist thought based upon a reductionistic view which equates it with vulgar economic determinism.[4] Some criticism of Marxist applications to media studies has been warranted, as in the writings of the Frankfurt school, where theorists argue that popular media texts contain insidious messages which directly reinforce the economic and political status quo (Adorno, 1991). The media - or as the Frankfurt school terms them, the 'culture industries' - serve as powerful instruments of social control in the hands of the ruling political and economic groups, they assert. While once

[4] For an example of one such dismissal, see Lull (1995). The dismissal of Marx by some U.S. scholars is also due in part to the U.S. resistance to analyses of class as well as to the deeply different understandings of 'progress' inherent in Marx's thought (as opposed to more liberal-pluralist Utopian visions, such as that of John Dewey).

culture had cultivated individuality, these theorists argue, through the popular media industries it has come to promote homogeniety and conformity due to the increased rationalization and bureaucratization of life in advanced Western societies.

The critical theorists, as these scholars are called, have been roundly criticized for their elitist approach to 'culture,' as well as for the way in which their theory approaches 'culture industries' monolithically (See, e.g., Storey, 1993). Yet theirs is but one, rather early, example of a critical approach to the question of determinism. The most explicit reference to economic determinism in Marx's original thought actually occurs in the writings of Friedrich Engels (1977):

> We make our own history, but, first of all, under very definite assumptions and conditions. Among these the economic ones are ultimately decisive. But the political ones, etc., and indeed even the traditions which haunt human minds also play a part, although not the decisive one (p. 76).

Yet even here one can see that Marx and Engels are not arguing that the form of culture can be *explained* by the organization of its economic resources. Rather, they note that the form of culture is produced by an interaction of this organization with other societal institutions and the culture's individual members. Individuals still have agency, therefore, although their ability to act is constrained by and occurs within frameworks over which they do not have control.

The critical/cultural approach to media studies owes much to Gramsci's foregrounding of the consciousness of the human subject, particularly as popularized by Hall, as noted earlier. In his influential article 'The Rediscovery of Ideology,' Hall (1982) notes that in the 1960s, scholars in media studies - in particular, those researching the 'limited

effects' of the media - came to recognize that rather than unproblematically influencing individuals directly, the media reinforce the values and norms which had already acheived consensual acceptance. Yet, Hall notes, the question then had to be asked was

> whether the consensus did indeed spontaneously simply arise or whether it was the result of a complex process of social construction and legitimation...For if the media were not simply reflective or 'expressive' of an already achieved consensus, but instead tended to reproduce those very definitions of the situation which favoured and legitimated the existing structure of things, then what had seemed at first as merely a reinforcing role had now to be receptualised in terms of the media's role in the process of consensus formation (pp. 63-64).

Thus, Hall argues that the media play a key role in the maintenance of cultural power systems through their ability to naturalize and legitimize some viewpoints over others, which is accomplished discursively, as "all discourses entailed certain definite premises about the world (p. 80)." Hall notes:

> Discourse, in short, had the effect of sustaining certain 'closures,' of establishing certain systems of equivalence between what could be assumed about the world and what could be said to be true...In this sense, Althusser was subsequently to argue that ideology, as opposed to science, moved constantly within a closed circle, producing, not knowledge, but a recognition of the things we already knew (p. 75).

While the power to define 'reality' is always contested, it cannot be simply understood as a class struggle, Hall asserts. He notes that the terms of ideology do not 'flow' from class positions, as ideological consensus is achieved through both conscious and unconscious thought. This is why, he asserts, individual broadcasters "speak" the dominant discourse, whether they are aware of this or not. As he notes, "ideology is a function of the discourse and of the logic of social processes, rather than an intention of the agent (p. 88)."

Critical/Cultural Studies: Identity as Positioned

What is it that makes cultural studies "critical?" Moores (1995) argues that it may be called 'critical' in the sense that the research

> takes extremely seriously the interpetations of media constructed by consumers in their everyday routines. At the same time, it is not afraid to interrogate and situate their spoken accounts (p. 5).

Thus 'critical' research implies an analytical task in relation to the ethnographic perspective; one does not simply categorize the functions of an individual's (or family's, or social group's) media practices, but must understand them in relation to other, larger contexts. Ang (1996) embraces a similar definition, although she also adds that critical research, in its distrust of positivist empiricism, must adopt

> a self-reflexive perspective, one that is, first, conscious of the social and discursive nature of any research practice, and, second, takes seriously the Foucauldian reminder that the production of knowledge is always bound up in a network of power relations (p. 36).

Thus in consideration of the larger context, one must explore both the power relations structuring the subject's position as well as that of the researcher. The term 'critical,' Ang further points out, also signals a link with both the Marxist and semiotic critiques.

When one compares the roots and aims of research within the critical/cultural tradition with that of the 'uses and gratifications' and American cultural studies traditions, then, it is difficult to conceive of how the two may "converge," as Rosengren (1983) has suggested (See also Linderman, 1997). As Ang (1995) writes of these differences:

> The aim of cultural studies is not a matter of dissecting 'audience activity' in ever more refined variables and categories so that we can ultimately have a complete and generalizable formal 'map' of all

> dimensions of 'audience activity'(which seems to be the drive behind the uses and gratifications project). Rather, the aim, as I see it, is to arrive at a more historicized and contextualized insight into the ways in which 'audience activity' is articulated within and by a complex set of social, political, economic and cultural forces. In other words, what is at stake is not the understanding of 'audience activity' as such as an isolated and isolatable object of research, but the embeddedness of 'audience activity' in a complex network of ongoing cultural practices and relationships (p. 42).

In an interesting paradox, therefore, I argue that critical cultural studies is both more ambitious and more modest than uses and gratifications research. It is more ambitious in that it always requires us to take the complex workings of historical, political, economic and other social contexts into consideration, and as such it demands analysis of an unwieldy amount of information. Yet it is also more modest in that its aim is not to construct a totalizing and finite picture of the world, but to construct a convincing yet temporary narrative that points toward an increased understanding of the workings of power in society.[5]

The tradition of media research associated with British cultural studies, as noted in the introduction, dates to a turn toward cultural issues within literary theory and later, an incorporation of psychoanalysis into film theory and an inclusion of concerns raised both in neo-Marxist and semiotic theory. Even with its structural/societal focus, however, there has been an influence from the field of psychology in the developments here. In the film theory associated with the British journal *Screen,* for example, Metz builds upon Lacan's theory of the "mirror-stage."[6] As Flitterman-Lewis (1992) explains:

Just as the infant sees in the mirror an ideal image of itself, the film

[5] Ang (1995) cites Grossberg (1987) as a source of an argument for the "modesty" of critical/cultural studies.

> viewer sees on the movie screen larger-than-life, idealized characters with whom s/he is encouraged to identify...This early process of ego construction, in which the viewing subject finds an identity by absorbing an image in a mirror, is one of the founding concepts in the psychoanalytic theory of cinema spectatorship... (p. 214).

Borrowing from Freudian psychoanalytic approaches, therefore, film theory addresses itself to a paradox in media consumption: by identifying with those on the screen, our own egos are temporarily eclipsed, yet simultaneously reinforced. This happens, Metz (1982) argues, because the author of the film is hidden and the spectator is invited by the text to identify with the subject. Yet this process of identification has been problematized, through both feminism's and postcolonialism's critiques of the naturalized masculine, Anglo and Euro-centric subject position. Feminist film theorists such as Mulvey (1975), Kuhn (1982), and de Lauretis (1984) note the way in which film addreses the subject (the audience member) as male, voyeuristically adopting a "male gaze" onto females who are objects of this gaze. Similarly, other theorists have noted the naturalization of a European/U.S. "whiteness" in the gaze of film texts. Shohat and Stam (1994), for example, cite Franz Fanon (1967) in *Black Skin, White Masks*, to point out the different processes of identification based upon the location of the viewer within the wider cultural sphere:

> Attend showings of a Tarzan film in the Antilles and in Europe. In the Antilles, the young Negro identifies himself *de facto* with Tarzan against the Negroes. This is much more difficult for him in a European theatre, for the rest of the audience, which is white, automatically identifies him with the savages on the screen (cited in Shohat and Stam, pp. 152-153).

Hall (1989), citing the same text, argues that identity is constructed in

representation to the extent that it can make some audience members see themselves as 'Other.' Employing Althusser to challenge the Lacanian interpretation embedded in traditional film theory, he writes:

> I have been trying to speak of identity as constituted, not outside but *within* representation; and hence of cinema, not as a second-order mirror held up to reflect what already exists, but as that form of representation which is able to constitute us as new kinds of subjects, and thereby enable us to discover who we are (p. 80).

In this example, we can see the influence of the Althusserian notion of ideology as that which creates subjects and gives them a position within the larger society as they are 'hailed' through interpellation - the process in which we are made to believe that we are unique beings yet are actually held in our place. Turner (1992) suggests that while Althusserian thoughts on the subject were key to media audience analysis in the 1970s and 1980s, his truncated analysis of the relation of the subject and social and cultural change ultimately led to an adoption of the more nuanced analyses of Gramsci.

Later adaptations of *Screen* theory were more sensitive to the notions of the 'active audience' which was arguably introduced in Hall's (1980) influential essay, "Encoding/Decoding." While the essay is seen as a turning point from the analyses of "meanings" encoded into media texts toward an interrogation of the reception of those texts, it also represented a significant shift in terms of the receiver/text relationship. As Turner (1992) notes:

> Where the earlier notion of the "effects" of the media localized the meaning (and the effect) of the message in the individual reader, the encoding/decoding model defined media texts as moments when the larger social and political structures within the culture are exposed for analysis (p. 94).

This shift necessarily influences the understandings of identity, as well. As Hall argues, the individual receiver's interpretation of television texts must be understood within the context of larger cultural discourses, the visual and oral conventions of the media, and the cultural status and function of the technology itself.

Early writings on identity in the critical/cultural tradition focused on the formation and maintenance of subcultural identity, privileging class-based analysis to the exclusion of virtually all other categories, as noted earlier (See, e.g., Hall & Jefferson 1976). During the 1970s and 1980s, however, scholars in Britain, like their counterparts in the U.S., responded to the challenges of feminist theory, incorporating gender, and later race, into their analyses as well, thus addressing identity and its relation to "difference." While it has been tempting for some critics to dismiss all work in this area as a dissolution into individualism and a celebration of "identity politics," I argue that this is not universally true in cultural studies approaches. This is because the addition of concerns of race, gender, and other inequities has not resulted in an "add women (or persons of color) and stir" approach, but has encouraged a reconceptualization of the very questions to which cultural studies addresses itself.

As I noted previously, there are signs of an emerging critical discourse in cultural studies which addresses issues of identity without an eclipsing of issues of power. Kellner's (1992) argument that postmodern television is not "depthless," as Jameson has argued, is one example. Kellner demonstrates that

> both the images and the narratives [in so-called 'postmodern' texts] are saturated with ideology and polysemic meanings, and...ideology

> critique continues to be an important and indispensable weapon in our critical arsenal (p. 147).

Thus Kellner reclaims 'postmodern' texts as sites for analysis, exploring the ideological themes in advertisements and popular television programs to demonstrate the fact that postmodernism's "identity crises" are rooted in and in some instances inseparable from the concerns of Modernity.

Additionally, Morley and Robins (1995) incorporate insights from poststructuralism, postcolonialism, and Marxist theory in their recent work, *Spaces of Identity*. They argue that Europe, the context of much media research, is not only a physical location but also an idea shaped by its relationship to the spaces it has defined itself against: notably, its colonies and its enemies. This work and others like it may serve to push cultural studies away from its focus on the audience and toward issues of the global distribution of communication technologies and their texts. Such widespread distribution does raise questions concerning the role of various places which are receivers of media texts who have limited means other than at the site of the audience to serve as producers of 'texts.' Further, the study of such issues raises questions of the role of the researcher and the issue of "talking about" rather than actually hearing from voices outside the Western-dominated academic discourse, not only in terms of informants for academic analyses but also to challenge the analyses themselves. As bell hooks (1990) notes, for example, while cultural studies opens space for discussions of imperialism, it is "disheartening" to see so little acknowledgment of the radical black discourse of both African-American men and women that shaped the discussion. She notes in a critique of the exclusive list of authors in the

progressive text *Writing Culture:*

> If white male scholars support, encourage, and even initiate theoretical interventions without opening the space of interrogation so that it is inclusive, their gestures of change appear to be ways of holding onto positions of power and authority in a manner that maintains structures of domination based on race, gender, and class (p. 131).

While the critiques of media literature which overemphasize the emancipatory potential of media texts are warranted, I believe critical/cultural studies has maintained its political commitments to a changed society as evidenced by many voices, such as Hall (1996), Morley and Robins (1995), Peck (1994), and Stabile (1995). These questions of research focus and the political nature of research itself are crucial to the study of media and identity - and are questions that will not be easily resolved. They point to the fact that critical/cultural studies is not automatically in a redemptive position; as Lopez (1992) argues, for example, it is the responsibility of those scholars working with subaltern groups not to render the practices of the oppressed visible to those who dominate, but "to make the operations of capitalism and patriarchy more transparent to the oppressed groups," which is a much more challenging proposition (cited in Katz, 1994, p. 71).

The reconceptualization of critical/cultural studies in relation to postcolonial theory has had the advantage of bringing together the concerns of Marxist critiques with the "lived" experiences of those who, due to political-economic arrangements of nation-states and the global effects of capital, acknowledge, as Williams and Chrisman (1994) argue, that the end of colonialism has not stemmed the "globalising spread of imperialism." Contrary to the other approaches to media and identity, the postcolonial approach holds the potential of foregrounding the

importance of power relations in a relational concept of identity, as Nagar (forthcoming) notes:

> they (power relations) are inscribed in dominant discourses, institutional structures and practices which rigidify certain identities and communities and de-emphasize others, thus defining rules of exclusion and inclusion...At the same time, however, communal organizations, socio-political, and discursive processes have operated over time to impart a structure and continuity to these categories, and to strengthen and reinforce people's identities as members of these communities (p. 6).

It is clear that in this perspective, the notion of ego as the Althusserian "always-already" subject is maintained yet also understood as shaped by a multitude of both material and ideological forces - thus both challenging the "centered" self of earlier approaches and materially grounding the ideological process in specific contexts of "lived" experience.

Unfortunately, when postcolonial theories are brought to bear on television and other mediated experiences (and even when they refer to their method as "ethnography"), they are often answering questions within the psychoanalytical framework that has been important in film theory's past and are less attentive to empirical questions of social structural concern (See, e.g., Shohat, 1993). However, some texts have examined the more traditionally 'culturalist' question of identity-construction at the site of audiences, texts, and larger cultural/historical forces, notably Naficy's (1993) recent work on exile communities in the U.S.. In this book, he explores

> the power of the media to enhance and consolidate subcultural identities based on location, ethnicity, race, class, religion, politics, language, and nationality. It also shows that the relationship between mainstream culture and subcultures is fraught with ambivalence and contestation on the one hand and enrichment and assimilation on the other (p. xvi).

With this example in mind, I argue that the question of the relationship of media and identity holds the potential to shed light on such issues theorized yet not empirically explored in much of postmodernist discourse: of multiple positionalities, hybridity, marginalization, and the decentred self. All of these issues not only suggest the need for the exploration of cultural practice and the histories which brought such differential/inequitable relationships into being, but a need to understand more fully the role of the media in this process of creating and maintaining - or perhaps, challenging - such situations. Such a contextualized approach is consistent with the foundational writings in cultural studies, but also with the Marxist and neo-Marxist assumptions about inequal power distribution and the means by which it is kept in place.

My own concept of identity grows out of this critical/cultural studies tradition. As I noted in the introductory chapter, I take identity to be not simply the sum of labels that describe an individual's demographics, nor the descriptor which refers to the "core" or "essence" of one's self, but as an elusive analytical category which may be helpfully employed to explore the interrelationship of the individual and his or her social and material environment. Identity, as it is constructed in narrative, foregrounds the individual's perspective on that relationship, yet the individual must draw upon existing categories to either describe or differentiate. Thus identity is constituted in the symbolic marking of boundaries, as the individual subject *articulates or claims* certain discourses as his/her own and thus expresses a perceived difference between self and other. Further, while the identity narrative itself may be

seen as temporary and constructed for the specific audience listening at the time, the frameworks of discourse upon which it draws are lodged culturally and historically. When a person constructs a narrative of identity, therefore, he or she is not simply speaking the subject position of the discourse as he or she is 'hailed' in it, but is actively claiming the narrative and investing in it. The narrative becomes an *articulation* of identity, to use Hall's (1996) term.

The problem with religion

In this section, I would like to explore the "blind spot" of religion within studies of media and identity by again relating the current situation again to its intellectual origins. As noted earlier, the thought of Freud and Marx, as well as the scientific, behavioral, and finally cultural studies traditions shaped by these and other thinkers, have each contributed to the truncated ways in which religion has been explored or dismissed in media research. Yet each also offers clues as to how religion might be reconceived given the current climate of research into media and identity.

Perhaps it may be true that, as Hoover and Venturelli (1996) have asserted, Marx himself was less hostile to religion than his followers have assumed. These authors have re-read Marx's famous statement in which he described religion as "the opium of the masses," arguing that while Marx "obviously conceived of religion as the illusory consolation of those who are in fact denied the possibility of real justice or happiness, he did not identify religion itself as the source of oppression or exploitation (p. 255)." However, there is no denying the fact that Marxist thought has largely been understood to be focused on concerns of the *material* as

opposed to the *ideal, supernatural* or *non-rational* spheres. This in turn has led to a tendency to dismiss religion as marginally important or to see it instead as a powerful force of ideological justification for the status quo. While we might wish to debate the usefulness of the material/ideal or rational/non-rational dichotomies, their influence on the thinkings of Marxist scholars has been enormous. A result has been the delegitimation of the realms of emotion, pleasure, and leisure in addition to religion - in short, anything not directly related to the relations emerging in the material production of goods or services for profit. Unfortunately, while critical/cultural studies has reclaimed each of the other categories, it has remained uninterested in the religious. One possible explanation for this emerges in a footnote to an essay on popular culture by Mukerji and Schudson (1991). These authors note that cultural studies is interested in the meaning-making practices surrounding the "low art" of the people, as opposed to the "high art" which members of the upper classes employ to maintain their cultural distinctiveness. In this way the authors dismiss religion as serving an ideological function similar to "high art," which may be one assumption governing its oversight throughout critical/cultural studies works.

Religion for Freud (1961) is in some senses similar to the analysis of Marx, although with a different valence: it succeeds in keeping "these dangerous masses..held down (p. 50)." In *The Future of an Ilusion,* Freud argues that God is a projection of the needs humans have for some understanding of the chaotic realities of nature. Like Oedipus, we relate to God as children:

> When the growing individual finds that he is destined to remain a child forever, that he can never do without protection against strange

> powers, he lends those powers the features belonging to the figure of his father; he creates for himself the gods whom he dreads, whom he seeks to propitiate, and whom he nevertheless entrusts with his own protection (P. 30).

For Freud, then, religion serves several functions for individuals: it provides a salve against the fears of nature; it reconciles people to the inevitability of death, and it provides some compensation for the sufferings of this life. The latter, in particular, serves to hold the rules of a society in place, for individuals must have some motivation for repressing their asocial instincts.

Agreeing with Marx, Freud argues that religion's influence will continue to decrease as scientific knowledge becomes more widespread. Unlike religion, scientific claims can be subjected to "proof," and thus the latter, in Freud's mind, is the superior system. We have already noted the connections between Freudian thought and the rise of a focus on the individual in the U.S. In this instance, Freud's thought also dovetails with the increasing trust placed in the dominant scientific worldview later associated with Parsons and Popper, which emphasized the strengths of logical positivism. It is perhaps ironic, then, that the psychoanalysis rooted in Freudian thought has taken on some characteristics of a totalizing world view similar to a religion. In fact, Michael Jacobs (1992), a biographer of Freud's, suggested that "psychoanalytical literature has replaced the Bible as the place to which people turn for an explanation of their suffering and a source of consolation (p. 121)." Indeed, I argue that much of contemporary religious discourse has adopted a therapeutic tone rooted in the psychoanalytic model of mental and emotional health (See, e.g., Roof, 1982).

Media scholarship within both Marxist and liberal-pluralist traditions has been buttressed in its tendency to overlook religion by the argument for secularization outlined in the introductory chapter. In brief, this argument asserts the rationalist Enlightenment idea that as societies and cultures become more knowledgeable of the scientific realities of the universe, religion will become increasingly unneccessary and hence will decline. Current proponents of the secularization hypothesis point to the decline in attendance among the formerly hegemonic mainline Protestant institutions as "proof" of this assertion. These scholars tend to overlook the growth of various fundamentalisms, new variants of Roman Catholicism, and the growing Muslim population in the U.S. (not to mention the rise of various "alternative" religions), all of which would contradict a simple instance of decline (Casanova, 1994). As noted in the introduction, the persistence of practices of popular rather than institutionally-based religion mitigates in favor of an inclusion of the analysis of religion, in particular as it emerges in the narratives of media meaning-making practices. The last section of this chapter will illustrate how this task might provide interesting insights into the development of a deeper understanding of "meaning" within critical/cultural studies audiencing analyses.

Joe and Amber Dearborn: A Postmodern Family

Joe and Amber are perhaps best described as a "postmodern family,"[7] as disruptions due to employment and domestic changes have figured prominently in their lives. Joe "won" 16-year-old Amber in a

[7] Stacey (1991) uses this term to describe the disruptions wrought in domestic life as a result of the changing economic order.

bitter custody battle after Amber's relationship with her mother Leanne deteriorated once Leanne moved then-15-year-old Amber and her brother Tim into her then-boyfriend, now-husband Trevor's home (where Amber lived for less than a year). While Tim stays with Joe and Amber every other weekend, Amber's visits to the home of Leanne and Trevor, with whom Tim and a stepsister Nan live, are far less frequent. Surprisingly, despite the contentious relationship between Joe and Leanne, Leanne's family still participates regularly in Dearborn family gatherings, and thus Amber continues to have an extended family of support.

Amber's move to her father's apartment further destabilized Joe's already-fragile financial situation. He has gone through frequent layoffs, once as a truck driver and now in his job as a heavy equipment mechanic, a position in which he estimates earnings of $30,000 a year.[8] The move also represented a financial change for Amber: after having been raised in Westville, a lower middle class area of the city with a large Hispanic population, she had moved to Trevor's home in Orchard Park, an upper middle class and primarily caucasian area. It is not surprising, then, that a relationship between geographical location, class, gender, and race converge in Amber's discussions of her identity-negotiations. Amber identifies herself as "Chicana," although her father corrects her, noting that he is part Native American and while they both grew up in Hispanic areas of the city, they have no Mexican heritage. She calls her friends her "Chicana sisters" and also claims Hispanic heritage for them (although Joe is skeptical about this, also). When she speaks of her Chicano/Chicana friends, Amber adopts an appropriate accent as if to signify aurally her

[8] They have since moved again to a trailer Joe is struggling to purchase.

connection with the culture.[9]

Amber differentiates herself from the "Orchard Park attitude," noting that she prefers her life "on the edge," where you must "do all you can to live:"

> Lynn: How come you didn't get the Orchard Park attitude?
> Amber: 'Cause I was more rebellious against it (than her brother was). Against living down there and stuff. If I ever saw it happening, 'cause I seen it happen to other people, and every time I look at them, it'd scare me even more and more, so. No way. I got into all that stuff (drugs, trouble, etc.), and that's kinda what kept me alive from turning - UGH! - into a scary person.

Amber sees her move from Westville to Trevor's home in Orchard Park as what Turner (1974) called a "liminal moment," a time in which she was forced to redefine herself because of her immersion in the new social and economic context of her more scholastically rigorous new high school and her mother's imminent remarriage. This change emerges in her narrative in discussions of her peers, behaviors, performance in school, and even music choices. It also surfaces in her conversations about religion.

Both Joe and Amber talk about religion as an important aspect of their lives, and were regular attenders and members of an American Baptist church during the time of the divorce and during Amber's move to Joe's. They now infrequently attend this church, which Joe's mother also attends, and where Amber has been active in leading the young children's programs in the past. Amber praises the church for the fact that its members treat her "like an adult" and show her "respect." They both consider their religious beliefs conservative. During one visit, Amber showed me a picture she had drawn based upon an illustration in a

[9] I later learn that she is using "Chicano" as a code term for her association with the Crips gang, of which her father is not aware.

children's Bible titled, "The Princess finds Moses." She noted, "I always use the Bible. It's really good." Other than this, however, there was little in their speech or in the artifacts in their home to indicate the importance of the Bible or other traditional religious paraphenalia in their everyday lives.

Joe had been raised United Methodist but his family became Baptist during the Vietnam war, when, Joe says, "the young pastor we had in there was preaching more politics than religion." They then joined a Baptist church as a family, but Joe "became disenchanted" with that church because of its hypocrisy:

> There was a lot of inner church politics, a lot of hypocrites, really. On Sunday, they'd be 'hey brother,' like that, but come Monday they wouldn't talk to you. A lot of holier than thou people who'd look down their nose at you, but at night they were doing worse things than I was doing.

Joe then quit attending church for many years, but six years ago began attending a Baptist church his mother and stepfather joined a few years earlier. Although there are very few people his age (Joe is 42), and even fewer that are Amber's, they both describe it as "like family."

Both Joe and Amber have told me of paranormal experiences they have had, although neither of them see many inconsistencies (or connections) between this and their conservative Baptist faith. Amber describes her first paranormal experience, which occurred when she was young, as a seemingly pleasant and benign interaction with her grandfather who had died violently in a fire. Her description of the experience has some similarities to Joe's description of his own paranormal experience, which occurred when he was a teen after his father died of a heart attack while performing on stage. He describes the

single occurrence after noting his inability to deal with his father's death at the time:

> Lynn: Did it hit you later?
> Joe: It hit me later, yeah, it did. (pause) Then there - you probably think - this is really silly. (pause) Ah, I don't know if I believe it any more. (pause). Two weeks after his death I got the room that he used to have, he and my mom used to have. It's a straight shot, I can look out that bedroom door, and I can see the back door, straight shot through the entire house. (pause) And. About 2 o'clock, 2:30 in the morning, you know I used to wake up and hear my dad come in the back door (after performing with local bands). And I woke up and looked up and here - (slight laugh) - as God is my witness, I swear I seen this. There was this black and white image of my father there. And he doesn't open the door, he comes *through* the back door. And he's tripping through, and he's walking. And it's a full body, black and white - and I'm awake. I am, so help me God, I am awake. I didn't dream it. I freaked, I watched right until he got to the beginning of the hallway, which is about 15 feet from my room, and I freaked, I just pulled the covers over my head. Looking back in retrospect, I really don't think the man would've hurt me. I think he was just so used to coming home he was coming home.
> ...
>
> Lynn: Did he see you, or was there any kind of connection?
> Joe: It was like you projected an image onto a cloud, like a holographic image. It wasn't like he was looking at me, it was like he just - had a really weird look in his eye, this distant look. Like, you know it is when someone's daydreaming, and they're looking at you, but they're not really looking at you, they're looking through you. That's the way he looked. It's like he just was walking like this, not really looking at me. I was twelve years old, I mean, shit. I was a horror fanatic. And here it is in living color and I'm like, 'well, TV's one thing, this is real life. We're outta here.'

Joe notes that it was the only time he ever saw this apparition of his father. Amber, on the other hand, described several paranormal experiences of her own, none of which echoed Joe's fear to quite the same extent. She also approached the topic much differently, seeming proud of and excited by her experiences while Joe was tentative in describing his. This is not surprising, given that he might have expected me as a

researcher to be skeptical of their truth claims and he might have feared that such tales might have even cost him legitmacy in my eyes. Amber, on the other hand, presented them as more evidence of the dramatic and vivid life she leads, which was a major theme in her self-presentation to me. What is interesting here is the connection between these experiences and other discourses, as I will discuss. One example is the discourse of fright and "evil" which emerges in Amber's description of a different yet similar type of experience during the liminal move to Orchard Park:

> Lynn: Have you seen anything recently, like since you've moved here?
> Amber: Not recently. But back at the old house, I saw it all the time. It was great.
> Lynn: And then did it stop when you went to Orchard Park?
> Amber: Yeah. I think something - when I moved to Orchard Park, I knew something was there but it wasn't a good feeling. I think it might have followed me from my neighbor's house. I used to mow his lawn all the time, and he was a trucker, and no one was in our neighborhood, and I was mowing his lawn and was almost done and I was just talking to myself and all of a sudden the lawn mower just died. It was full of gasoline, and all of a sudden I heard this really evil laughing from all around. Nobody was in the neighborhood except for me and my brother. And it sounded really evil, and then it kinda followed me down to Orchard Park. And I saw a figure...it wasn't like a full body, it was kinda like an aura around a face...Really gruesome teeth, no nose, and that's the aura right there (pointing to a picture she'd drawn of it earlier). And I saw it at school and when I went home, and in my nightmares, and I never got a full night's rest because of that.
> Lynn: And so that started before you moved to Orchard Park and then was there?
> Amber: Well, I didn't see too much of it but I saw it mainly when I was in Orchard Park. I'd go outside and it would follow me. I felt I could feel its vicious paws go through my body. It felt like I had been torn in half and then put back together again. It was really scary. It followed me a little at night, and I haven't seen it since. (pause). I don't know. Some people don't believe in ghosts, but I do.

Employing a psychoanalytic framework, one might draw connections between Amber's unhappiness and alienation in Orchard Park, her

parents' divorce, and her experience of this evil manifestation which tore her "in half and then put [her] back together again." Yet what is particularly interesting to me is that she talks about this experience openly, even mentioning it to her Sunday School teacher. In fact, when Amber told her Sunday school teacher about her experience, the church leader equated it with a manifestation of "the devil," an explanation that seemed acceptable to Amber. Joe, on the other hand, was somewhat shocked to learn that Amber had mentioned it to the church leader. He himself has not talked about his own experience with anyone in the church, although he has told a number of friends about it. This underscores, again, his sense that discussions of the paranormal are illegitimate in public discourse and in the context of "traditional" religion. He attempts to control when and in which circumstances he will participate in the discourse about the paranormal.

Amber and Joe both like to watch documentaries and other programs about paranormal experiences, grouping these "strange phenomena" with "life after death experiences," "faith healing," and "ghost stories." Joe's use of the latter term is not meant to trivialize the stories, as he argues for at least the possibility that they might be "true." He and Amber both claim expertise in being able to tell an accurate portrayal of the experiences from one produced by someone who doesn't "have a clue of what they're talking about," as Joe says. In this sense, then, they are able to be 'hailed' by the discourse of these programs and yet also articulate their identification with the content as "experts" on the paranormal. Joe mentioned as particularly appealing the program *In Search Of* , which had been narrated by Leonard Nimoy. They distinguish

these shows from the "science fiction" of *The X-Files* and *Outer Limits*, which they do not watch. Yet the two activities that they mentioned they do together regularly as a family are attending church every week and watching *Star Trek* every night at 10:30 P.M..

Their interest in the paranormal, and its popular portrayal in the media, does bleed into their religious beliefs, as can be seen in this discussion of *Star Trek* , in which a statement of cosmology emerges:

> Lynn: Do you think there's any connection between your paranormal experiences or religion and your liking of *Star Trek* shows? Because some people seem to think they're sort of religious, or uplifting, or whatever.
> Joe: I wouldn't say so. Even my pastor and I have gotten into debates over *Star Trek*, and Gene Rodenberry in particular. Because he, my pastor believes that *Star Trek* is along the lines of the cults out there that are saying that Jesus Christ was actually an alien that landed on this planet. And I'm goin'...'Well, do you know he wasn't?' And I said, 'y'know, that is one man's point of view'. I said, 'there's too many things in this world that nobody can explain away, like the drawing on the pyramids with the ancient Egyptians, drawing what looks like a modern-day space suit. And too many other variables, y'know?' I said, 'all I'm saying is, *what if*?' And he just, you can see the veins start to come out (motions to his neck), but him and I, we go at it. I don't think, as far as *Star Trek*, I'm just curious. I'm just a natural curious person. I truly believe that we don't know a fraction of the things that go on in this world. So, I don't think *Star Trek* is something that really pulls any weight on paranormal existence, I just think that it might offer some ideas on why it takes place. Because, *Star Trek*, along with a coupla other shows, are based on - everything that is done on *Star Trek*, they may stretch the hell out of it, but if you bring it back down, there is a basis in scientific fact about a lot of things they do, as far as, like, the transporter. The particle beams. The carrier waves. There are carrier waves that we use in radio. Particle beams, we've got lasers now, they're particle beams. Yes, they may stretch it until it's almost unrecognizable, but everything they - and y'know, what if life after death is nothing more than a different dimension? You go, from, say, this dimension to another? I'm not saying we do that, I'm just saying, *what if.*

This is a rich exerpt and worthy of several comments. First, I would like

to explore the relation of *Star Trek* to myth, acknowledging that several other scholars have noted this connection (See, e.g., Jindra, 1994; Amesley, 1989; Jenkins, 1988). One of the most helpful definitions of myth in media is offered by Silverstone (1981), who draws upon the writings of Cassirer, Eliade, Levi-Bruhl, Levi-Strauss, and Malinowski for this task. He defines the mythic as

> traditional stories and actions whose source is the persistent need to deny chaos and create order. It contributes to the security of social and cultural existence. The mythic is a world apart, but it is also close at hand. It acts as a bridge between the everyday and the transcendent, the known and the unknown, the sacred and the profane (p. 70)."

Let us look at this definition in greater detail. Silverstone refers to myths as *traditional* stories because they are conservative in a fundamental way. Yet to be effective, they must be flexible and open so as to incorporate new narrative and symbolic elements which harmonize emergent ideas in culture with their original structure and content. *Star Trek* sustains a traditional myth of the hero's quest for justice and for an egalitarian society while locating this quest in a futuristic setting. Silverstone notes that as *stories,* myths rely upon narrative structure, moving from uncertainty to certainty. The desire for resolution of uncertainty is one of the motivations of the mythic, and this resolution, even if symbolic, plays a functional role in culture as it contains the irrational. Because myths frame the irrational and the rational beliefs of a culture, they can become sites of contestation; they are *social* constructions. Those in the dominant positions of a society may be expected to reinforce the mythic in terms of their own vested interests, sometimes even laboring to secure alternative meanings for myths (Geertz, 1968). Myths are ideological, and as such, their meanings are never finally secured. Because they draw upon Utopic

desires through the emotions, they appeal to various members of culture for different reasons, and thus they are always open to redefinition. As a mass produced text, however, *Star Trek* does frame a certain set of beliefs, excluding some while highlighting others.

Jindra (1994) has argued that *Star Trek* offers a Utopian vision of the world in which "faith is placed in the power of the human mind, in humankind, and in science (p. 32)."[10] He argues that it offers a world view which functions much as religion or myth, noting the centrality of progress to this view. Yet in the case of Joe and Amber,we do not see a family whose world view mirrors that which is seen on *Star Trek.* Instead we see a family that is self-identified in the traditional religions, espousing a theology which may draw upon *Star Trek* but for which *Star Trek* is clearly one of several sources which also must include the various mediated explorations of paranormal phenomena. A discourse of progress never arose in my conversations with Joe and Amber. Both, in fact, were skeptical of *Star Trek*'s egalitarian themes, which Jindra suggests are widely accepted among the middle class, such as the possibility of societal economic equality or the likelihood that various racial/ethnic groups could coexist in harmony. This illustrates the ways in which they *fail* to be interpellated into these discourses of the "dominant" culture and thus assume a position of negotiation. Instead, Joe and Amber emphasize the authority of their experiences of the paranormal, along with, perhaps, an implied discourse of relationality and connection with their ancestors. While they did not claim a link between their experience and Native American background, their interpretation of their experience might be suggestive of this.

[10] For a similar argument see Selley (1990).

Actually, Joe articulates better "critical viewing skills" than several other families I have interviewed. For example, possibly due to his class position, he recognizes that the news media portray only a "certain point of view" and have a tendency not to highlight the views of "common people" like himself. He notes that often what is seen in the media must be supplemented by information from more informal channels, particularly regarding news about his low-income neighborhood. Further, when I asked Joe if it bothers him that sometimes the paranormal activities are sensationalized in the media, he responded emphatically, "No, 'cause that's *entertainment.*" Thus while he is not seeking information on paranormal experiences directly from the media, Joe notes that he is a "curious person" and approaches the media as introducing new possibilities.

Critical viewing skills cannot be equated with total agency at the site of the audience, however, for the audience member is still situated/subjected by discourse. Accepting the discourse of science, for example, Joe is willing to leave the possibilities introduced by mediated messages open until they can be "disproved." I argue that the self-identification of Joe and Amber as persons who have experienced the paranormal shape their selection of content that satiates their curiosity about paranormal experiences. These choices, in turn, expose them to new ideas of the paranormal, which then shapes their interpretation of their own experiences, their audience meaning-making practices, and their interpellation into "vernacular" religion. Interpreting*Star Trek* in light of the paranormal and mystical learnings gathered elsewhere by Joe and Amber, does not simply result in *negotiated* meanings, to use Hall's

(1980) term, but may be said to participate in a *regeneration* of meaning. I use this term to refer to the practice of adding meaning to a mediated text which then informs the larger system of beliefs, as illustrated in Joe's statement about Star Trek's scientifically plausible explanations for paranormal occurrences. This practice also differs slightly an *oppositional* reading in Hall's sense, for it is not completely distant from the original meaning; in fact, it is the original context of science fiction as presenting a scientifically plausible explanation that assists in his construction of a religious worldview. When he relates examples from *Star Trek* to explain to me his own views and thus construct within our conversation an element of his own religious identity, he is noting that the negotiated meaning contributes something to his larger believe system - even as he denies that this is happening.

Conclusion: Religion, Identity, and Meaning-Making

As I have noted, there is a tendency in the U.S. to bifucate religion in our analysis of it: either we understand it solely in terms of its institutional and confessional forms, or we abstract it from public life completely, viewing it as wholly privatised practice (Hoover, 1996). This diverts us from exploring the interactions between the ways in which religion is expressed in various public aspects of U.S. culture (including the mass media), and its expression in "private" life. As noted earlier, I agree with the many scholars who have contested the secularization hypothesis, arguing that rather than declining, religion is emerging in culture in new forms. While membership in institutional religion is declining, there are many - even within traditional faith groups - who

embrace religion in forms such as the New Age and nature religions and understand themselves as religious or spiritual yet autonomous in relation to religious institutions (Luckmann, 1991). Albanese (1993) notes the ways in which these alternative belief systems have become infused into public discourse, thereby changing not only religion but also signalling the legitimizing of the discourses of therapy and environmentalism just as the legitimacy of more traditional religion in public discourse has faded (See also Hoover,1995). It is not surprising, therefore, to find individuals situated in traditional religion who draw upon and may be understood as situated in relation to a variety of other complementary and contradictory discourses, as Joe and Amber are. The struggle over the legitimacy of the discourse of the paranormal in Joe's narratives highlights his own sense of agency regarding his ability to articulate his identity using a variety of discourses. This occurs even as the public definitions of the paranormal frame his own understanding of this experience.

It is also interesting to note that while Amber has no difficulty situating herself in paranormal discourse, she was not able to articulate a cohesive and synthetic cosmology at all, which I have found is not at all uncommon among my teen informants. This left her with rather contradictory explanations of her experience as "paranormal" and as a manifestation, in Christian discourse, of "the devil." The fact that the non-"evil" paranormal experiences continue to hold legitimacy for her on the basis of her experience (the legitimacy demonstrated in her willingness to talk about it) indicates the extent to which the Baptist explanation may not preempt the contradictory discourse of the paranormal. It is at least

interesting to note the fungibility of the boundaries between cosmology/theology, the traditional institution, and science fiction for both Amber and Joe. Clearly, while there might be an inconsistency, they are not living on a "map" which separates the spiritual experience from the "secular" realm of the media.

As I will describe more fully in a later chapter, Joe and Amber both situate themselves within a religious discourse that I label *mystical.* I found that a number of individuals employing this strategy, perhaps not coincidentally, saw themselves at the mercy of larger systems - both economic as well as spiritual - over which they had no control. I recognize that what I have described here, however, is a single case study and thus I am not attempting to make any claims for its generalizability. Yet the fact that these narratives emerged *at all* illustrates at least that a blurring of boundaries between religion and the sense-making practices of media consumption is possible as individuals negotiate with media texts and other public discourses in their identity-construction practices. In this chapter, I have demonstrated that as audience researchers, we are afforded a much richer understanding of meaning-making processes as we take religious identity as constructed through articulation in discourse into consideration. This analysis challenges liberal-pluralist notions of a unified, wholly autonomous and agentive subject through an examination of the choices individuals make within their limited options. It also challenges both the vulgar Marxist privileging of class as the sole determinant of meaning-making practices and demonstrates what is missed when religion is overlooked altogether, as in much of the research in American and critical/cultural studies. This chapter also demonstrates

that the goals of such an in-depth, critical analysis are quite different from the simple assertion that "audiences are active," while also revealing power relations that might be overlooked in a more psychoanalytical or developmental approach to identity. In the next chapter, I will delve more deeply into the methodological implications of this critical/cultural approach to identity.

Chapter Three

The Method of Critical Ethnography: Linking Reflexive Empirical Research with Critical Theory

In this chapter, I will discuss my definition of critical ethnography and its connection with the critical/cultural studies approach to identity outlined in the previous chapter. I will introduce each of the case studies of this project and will describe the methodologies of in-depth interviews, participant observations and peer-led discussion groups which formed the basis for my research. I will also describe the importance of reflexivity in my work, arguing for its importance in any ethnographic research and for its significance in the development of an understanding of identities constructed in discourse.

As noted in the previous chapter, my definition of "critical" research includes attention to the contexts that shape power relations, a self-reflexive approach, and a link with both Marxist and semiotic critiques. I argue that contemporary critical theory's contribution to social understanding, as Morrow (1994) has asserted, lies in its ability to overcome the polarization between *Verstehen* (the Weberian concept of understanding through observation and interpretation) and *Erklaren* (causal explanations based upon invariant laws, such as Marx's evolutionary approach to human history). Another way to say this is that critical theory attempts to synthesize the analyses of structure and agency, such as in Giddens' (1984) structuration theory or Boudieu's (1977)

description of the *habitus*, which refers to habitat, inhabitant, habitual practices, and social patterns to habits of thought.[1] Critical ethnography therefore suggests an analysis which begins with interpretation of empirical data gathered through observations, interviews, and informal discussions, yet moves toward placing the subjective reports of the research participant and the researcher's analysis within the larger, Marxist-informed framework of critical/cultural studies.

Ethnography in Media Studies

Communication research has taken an "ethnographic turn," as several scholars have now termed it, borrowing methodological insights from the fields of anthropology, feminist studies, and interpretive sociology (See, e.g., White, 1994; Moores, 1995). Methods specific to the field of media are still being discussed and developed, however, primarily because interpretive work is an expensive and time-consuming undertaking and, as it is still seen as a "minority" approach in the shadow of the dominant quantitative/positivist paradigm, generally receives less support and attention.

Although in U.S. media studies it is currently fashionable to claim that our qualitative methods draw upon cultural anthropology's approach to ethnography, the tendency to draw upon data gathered in a series of single, isolated interview situations as opposed to engaging in an extended

[1] Morrow suggests that it is perhaps most helpful to consider critical theorists as tending to fall somewhere on a continuum between interests in interpretation or causal analysis, and between interests in structure and agency. Morrow argues, for example, that while Habermas is a "strong" critical theorist for his interest in laws and structure, and hence is closer to Marxist and Weberian approaches, Giddens might be considered a "weak" critical theorist for his more empirical work which gives greater attention to agency. A similar dichotomy is noted in Jensen (1990). However, Jensen describes it within media studies as the overemphasis on the text and ideological impact (on the "strong" side, using Morrow's typology) versus the celebration of the "active audience" (the "weak" side) on the other.

period of living in a certain community actually places our practices much closer to those of interpretive sociology. In the latter field, it is equally common to conduct several (or even single) interviews and then move to an analysis of the data. In part, the shorter time in the field is justified by the researcher's supposed familiarity with the cultural context and language when research is done in the U.S. or in Europe; secondarily, it is justified in relation to a more "focused" approach to research questions (See, e.g. Lindlof, 1995).

Qualitative research into media audience practices has frequently involved interviewing persons about *specific media texts,* following Morley's (1980) classic analysis of Britain's television program *Nationwide.* I would therefore like to introduce the prominent studies of audience reception of specific media texts, which might include the works of Morley, Ang, Radway, Jensen, and a few others.

Although groundbreaking in its attempt to bring an empirical framework to Stuart Hall's (1980) influential "Encoding/Decoding" essay, Morley's study was criticized for its failure to consider the social settings in which viewing actually occurs, and for its selection of a specific program his research participants may not have chosen to watch on their own at all (See, e.g., Moores, 1995). Ang's (1995) study *Watching Dallas* has been criticized for some of the same reasons. Her unique focus on the practices of interpreting a popular U.S. television program outside the U.S. context, however, mitigated these criticisms to a great extent. Ang concluded that viewers make sense of Dallas when their own life experiences find some resonance with the "tragic structure of feeling" embedded in the text (p. 79). Thus in some ways her work contrasts with the fears of an

"Americanization" of the world through the increasingly prevalent satellite delivery of American television. This approach holds some similarities, then, to Latin American cultural studies, which challenge a blatant "cultural imperialism" approach to media reception.[2]

Another influential project was Radway's (1987) study of the domestic context of reception. Criticisms of this study have centered around the place of Radway's claims within the "politics of resistance," a particular strand of cultural studies literature I will describe more fully later. Corner, Richardson, and Fenton's (1990) *Nuclear Reactions* is another example of the approach which tends to privilege media texts. This study, which explores interpretative strategies of viewers of documentaries on nuclear power, outlines the ways in which work situations and previously held views shape interpretations of polemic material. They termed their work "ethnodiscursive" for its attention to the linguistic features of the respondents' talk. Jensen's (1986, 1995) studies of television news reception hold some similarities to this work, in that he also emphasizes the specific texts of news while exploring patterns of speech that emerge in his respondents' qualitative interviews as what he terms "super-themes."

A second approach to reception studies involves interviewing persons who are *members of a common subcultural group* to determine patterns in their media practices, such as in studies on the relations between subculture, aesthetic style, and media in the works by Willis (1977), Hebdige (1979), and in the edited volume by Hall and Jefferson (1976). McRobbie's (1991) work, which challenges the masculine

[2] This argument is elaborated in Murphy(1996). Martin-Barbero (1993) is perhaps the most renowned of Latin American cultural studies writers offering an alternative to the "cultural imperialism" approach.

assumptions within these analyses of subcultural groups, might also be considered here, as she explores media practices primarily through the categories of gender and age (youth). As each of these studies focuses specifically on adolescent subcultures, I will discuss them more fully in the subsequent chapter on adolescent audiences. Another important strand in this group of reception studies, however, has given attention to audience interpretations of representation and race or ethnicity, such as in the work by Jhally and Lewis (1992), and Naficy's (1993) work on television practices among the Iranian immigrant community in Los Angeles.

A third strand of audience research informing the current project involves those efforts that have foregrounded the *context of the home.* One notable study is the work of Silverstone and his colleagues (1991) in Britain, who "mapped" the uses of information and communication technologies in two homes. This study, as well as others of this type, have tended to emphasize television and its *use* for a number of reasons. Television, as Silverstone notes, is the "leading object" in what he calls the "moral economy of the household," as it serves to gather the family, structure the time of family members, and provide an important link between household and more public contexts. Hobson's (1982) interviews and observations of women (also in Britain) led to interpretations of meaning-making within the routines of housework. Her work, based in feminist critiques of the family, notes that for women, broadcast media are not associated with leisure but instead serve as companions and sources of information on the outside world which punctuate everyday housework. She relates this to larger issues of the legitimation of social practices, such as the tendency for women to see their own preferences for soap operas as

less legitimate than the "masculine" preferences for news and public affairs programs. In the U.S. context, Lull (1980, 1988, 1990) has engaged in extensive participant observation combined with some interviewing, also developing influential theories of media practices in the home. Lull's work is unique for its depth of participant observation work; most of the other studies in each of the paradigms discussed have relied primarily upon group and individual interviews, often not even in the homes of the research participants. Lull (1980) argues that television's functions within the home are both temporal and relational; it structures the daily flow of life as events are planned around the television schedule, yet also provide opportunities for both the avoidance or pursuit of intimacy between family members and "incessant opportunities for argument" as it becomes a site for domestic power struggles (p. 206). As Moores (1995) has pointed out, however, Lull's work on the family does not go far enough in its analysis of the power differentials structuring family media practices. Morley's (1992) more recent study, *Family Television,* addresses some of the same issues with a greater attention to explorations of how power differentials work themselves out in media practices. While insights gained from Lull's work are echoed in the findings of my own study, my own commitment to a grounding of issues within larger Marxist, feminist and third world critiques separates my work (and others of the critical tradition, such as Morley, Hobson and Ang) from Lull's largely functionalist and interpersonal framework.

Very little of this qualitative research has been directed toward families with teenagers, although such research methods are increasingly employed in studies of small children and their media use (See, e.g.,

Alexander and Munoz, 1984). Studies of teenagers have relied primarily on peer interviews and observations, an important approach to understanding the teen experience, yet not one that is usually contextualized with much reference to the domestic, familial context, as I will discuss in the following chapter.

Realism and the Critiques of Postmodernism

It is fair to say that the bulk of the qualitative research produced in media studies prior to the last decade - even that within the 'critical' tradition - was in the form of "realist tales," to use Van Maanen's (1988) phrase:

> These tales provide a rather direct, matter-of-fact portrait of a studied culture, unclouded by much concern for how the fieldworker produced such a portrait...The representation of social reality [is] seen as technically unproblematic once the facts [have] been unearthed (pp. 7, 19).

Van Maanen notes four characteristics of the "realist tale:" first and most importantly, the author is almost completely absent from the final report. The assumption is that what is ultimately reported represents what any person would have observed or surmised, had they been in the field instead of the researcher. Second, the researcher strives to demonstrate the "typicality" of the persons and situations observed and interviewed, thereby justifying their worth in relation to other, more positivist studies. Third, the researcher attempts to present "the natives' point of view," to use the famous phrase articulated by ethnography's father, anthropologist Bronislaw Malinowski (1922). This has come to include not only what the person says and does, but also, as Geertz (1973) has asserted, how the person makes sense of his or her actions and statements and how this fits

into the larger sense-making schema of the culture. Finally, realist tales, Van Maanen asserts, take for granted that their interpretation is the correct interpretation of the situation observed. These tales usually do not include evidence that might contradict their findings; the data is instead coded and categorized to support the ethnographer's conclusions.

In media studies, this 'realist' approach has frequently been paired with what might be called a "humanistic impulse." In her study of women reading romance novels, for example, Radway (1987) was particularly interested in "taking the women's self-understanding seriously," which she argues was a missing element in the feminist critiques of literary texts up until that point. Lull (1990) rather directly pits the benefits of ethnography's humanism against critical theory:

> Giving the audience a 'voice' has produced tangible, positive results. It has tempered the often pretentious and opaque writing associated with cultural studies, the frequent imposition of privileged interpretations of texts, a reliance on fashionable literature, and the dogmatic refusal to cite 'scientific' studies. Still, there are problems in the development of audience research in cultural studies. Most troubling is that we often hear too clearly the theorist's voice presented as if it were the audience's voice (p. 16).

Lull, Radway and others have celebrated "giving voice" to research participants, thus validating the subjective experiences of media use among those studied. This humanism, of course, has been disparaged by those critical theorists such as Garnham (1995), who seems convinced that by highlighting subjective experience, we lose our focus on structural constraints. Yet it has also been increasingly denounced by researchers who embrace ethnographic methods, particularly as a result of feminist critiques. Feminists, while wishing to "give voice" to women's experience, have also noted the need to place analyses of subjective

experience within larger frameworks of social constraint (usually those determined by gender; see, e.g., McRobbie & Gerber, 1976; Cirksena & Cuklanz, 1992). Some feminists working within a critical tradition, however, have sought to address the multiple forms of social constraint within U.S. society, as in Press's (1991) study of working- and middle-class women watching television. As Nagar (1997) argues:

> The privileging of dominant memory and public domain has largely excluded the experiences of all subalterns from standard academic discourse and treated them as a residual category since they are not the immediately visible participants in politics, trade, and matters of state. Through personal narratives of women and other oppressed peoples, oral and life historical research helps us understand the experiences of those who do not have access to means of publicity and the histories and geographies of whose feelings, thoughts, and actions get hidden behind the experience of the male middle class which incorrectly acquires universal significance (pp. 7-8).

Thus is it important to note that even the humanism of ethnographic research is not necessarily in conflict with the politics of critical thought. I find that by allowing my analysis to be informed, but not dictated, by Marxist and neo-Marxist thought, I am made more aware of the constraints under which subjectivities are reported - something that may be missed in the more humanistic reports. This then allows my research to contribute to the exposing of power relations within society, which I believe is an important goal of critical research and speaks to its aims of a more just society.

In addition to the challenges to its humanism, the "realist" approach to ethnography has come under considerable attack recently within anthropology, feminist studies, and postcolonial critiques, and these critiques are increasingly finding an influence in media ethnographies, as well. A particularly influential text is Clifford and

Marcus' (1986)*Writing Culture: The Politics and Poetics of Ethnography.* In this work, the authors identify "ethnographic realism" as "an ideology claiming transparency of representation and immediacy of experience (p. 2)." Clifford (1986) links this ideology to a fixed notion of culture and the presumed ability to scientifically describe it; instead, he argues, cultures, like the texts that describe them, are historical products:

> Cultures are not scientific "objects" (assuming such things exist, even in the natural sciences)...If culture is not an object to be described, neither is it a unified corpus of symbols and meanings that can be definitively interpreted. Culture is contested, temporal, and emergent. Representation and explanation - both by insiders and outsiders - is implicated in this emergence (pp. 18-19).

As a result of these and other critiques of a centralizing, essential notion of culture and its proposed description, issues of "marginality" and of "boundaries," of "hybridity" and "deterritorialization" have come to the fore in anthropology and in cultural studies more generally (See, e.g., Tsing, 1993; Appadurai, 1991; Abu-Lugod, 1991). Each of these issues points to the inherent limitations of any "grand theory" which might present (and *represent)* a plausible and complete model of culture and its workings. In this sense, cultural anthropological writings have dovetailed with the critiques of knowledge made by Foucault (1972), Derrida (1978), and Baudrillard (1978).

While Clifford and Marcus delegitimated the authority of the ethnographer who disappears in the "realist tale," they have been criticized for a failure to acknowledge the feminist and postcolonialist writings which have also contributed an understanding of the consequences of the construction of the "other" from a hegemonic, Anglo/European position.[3] Communication scholars Natrajan and

Parameswaran (1997) point out, for example, that Clifford and Marcus privilege the politics of *writing* ethnographies to such an extent that any other entrance of politics into consideration - for example, the power to write about the natives in the first place - is elided (See also Visweswaran, 1994; hooks, 1990; Behar, 1993).

To some extent, researchers embracing these critiques of ethnography have tended to eschew attention to issues of validity and reliability, claiming that such issues, as Lunt and Livingstone (1996) write,

> are inextricably linked to quantitative methods and so are irrelevant to qualitative work. Consequently, a researcher's attempt to take a distanced, autonomous view of data would be interpreted as a rhetorical stance that attempts to construct the researcher as expert by objectifying the participants as objects of research (p. 90).

While I agree that the key methodological issues of validity and reliability, operationalization, sampling, and even representativeness grow out of a quantitative research paradigm and therefore are less relevant to critical ethnographic research as the type of knowledge gathered in the latter is more modest and arises from different epistemological assumptions, the dominance of the former paradigm invites some grappling with these notions. I will therefore attempt to address some of these concerns in the section that follows.

With a few notable exceptions, all of the critiques of ethnography mentioned here have been slow in changing the practices of conducting and reporting audience research.[4] Instead, ethnographic media research has tended to advocate a "naturalist" as well as a realist approach in the belief that our best findings will result from the closest

[4] Two noteworthy exceptions, however, are Walkerdine (1986) and Lindlof and Grubb-Swetnam (1996), both of which have conducted and reported audience research reflexively.

approximation of the actual setting of media consumption were a researcher not there to observe these practices (Lincoln & Guba, 1985). Of course one always must wonder how possible it is to "erase" the researcher in such a way. As this is clearly impossible in the research situation, what interest is served in keeping him or her invisible in the final report? As my argument thus far suggests, such an erasure serves to reinforce the authoritative claims concerning the generation of scientific knowledge that require a neutral and objective researcher and setting. Ang (1996) makes a similar point:

> The comfortable assumption that it is the reliability and accuracy of the methodologies being used that will ascertain the validity of the outcomes of research, thereby reducing the researcher's responsibility to a technical matter, is rejected...the empirical is not the privileged domain where the answers should be sought. Answers - partial ones, to be sure, that is, both provisional and committed - are to be constructed, in the form of interpretation (p. 47).

In fact, many media researchers have risen in defense of the naturalist/realist method of media research (See, e.g., Lull, 1990). Moores (1995), for example, indirectly defends the writing style of ethnographic realism as he claims that reflexivity is not needed due to the "active critical capacities of readers" of audience ethnographies (p. 64). Moreover, Moores expresses the common concern that self-reflexive researchers who are themselves acknowledged fans may not maintain the "distance" required to "develop any kind of criticism," therefore risking a "populist acceptance or celebration" of the audience practices they are observing (p. 69). Yet does "distance" between oneself as a researcher and the practices one observes truly guarantee an objective position from which such criticism can be made? Reseachers of such common practices as media consumption should be particularly wary of this claim, for who among us

is not a television viewer? And by extension, then, how do we know when we have *enough* distance from our own experiences and practices to warrant a sufficiently critical interpretation? Because sense-making on the part of the researcher must always rely upon the knowledge we bring to the research context from either reading or personal experience, it seems to me that our *awareness* of our relation to that which we are observing and analyzing is particularly important. Acknowledging this awareness in our interpretations, then, may lead us to new insights into the nature, purpose, and politics of the knowledge being presented, which in turn shapes our understanding of the phenomena we study.

Furthermore, while critiquing any researcher's report, as Moores writes, is a widely-accepted professional practice, I argue that this does not address the basic epistemological challenge to the scientific construction of knowledge which underlies the critiques of ethnographic writing in Clifford and Marcus and in their critics. The questions raised here do not concern the accuracy of the particular reported research, as Moores seems to assume, but of the status of that knowledge itself: how it is constructed in the interpersonal relations between researcher and participant, and how the researcher's own position has influenced the research design, the researcher/researched relationship, and the "knowledge" that therefore results from the research project. Therefore we do not approach the research participant's statements as "raw data" that might be unproblematically analyzed but, to use the metaphor of the "linguistic turn," as *texts* which of necessity require interpretation.

I want to extend the argument for a self-reflexive project a bit further in the particular case of this research, which has foregrounded

issues of identity. As noted in the previous chapter, I assume that identity is never a "fixed" category (or sum of categories) that may be unproblematically interrogated but rather *it exists as it is enacted.* In the case of ethnographic research, identity is enacted in the behaviors observed and the narratives constructed for the audience of the researcher. Those behaviors and narratives are of course contingent upon how the research participant "reads" the situation, the researcher's intent, and the researcher's position *vis a vis* both the subject and the larger culture. In a very important sense, then, any analysis of identity-construction cannot be separated from the specificity of the researcher/researched relationship. The more insight afforded into the factors shaping that situation, the more likely we are to understand the knowledge produced. This does not require an academic's "life history," but it does underscore the fact that all knowledge is contingent, that all experience is interpreted from a subject position, and that the specific knowledge constructed in the interview process is, as Visweswaran (1994) argues, "itself determined by the relationship of the knower to the known (p. 48)." Reflexivity, therefore, subverts normative understandings of knowledge-gathering and information-processing; as Babcock (1980) argues,

> By confounding subject and object, seer, and seen, self and other, art and life - in short, by playing back and forth across terminal and categorical boundaries and playing with the very nature of human understanding - reflexive processes redirect thoughtful attention to the faulty or limited structures of thought, language, and society (p. 5)."

Reflexivity on the part of the researcher therefore challenges the "givenness" of data interpretation and the ontological and epistemological assumptions undergirding positivist research. Unfortunately, when media studies has engaged in such critiques it has tended to leave little

space for actual ethnographic investigation, as Bird (1992) notes. Murphy (1996) further illustrates the problem by noting that in a 1989 volume of the *Journal of Communication Inquiry* titled "Cultural studies, ethnography," each article is concerned with the politics of ethnography, yet *not one* actually illustrates the practice! (Murphy, 1996, p. 47). From these deficiencies, one might assume that ethnographic research is impossible or unfruitful in media studies. Yet I agree with the anthropologist Sherry Ortner (1984), who argues:

> It is our capacity, largely developed in fieldwork, to take the perspective of the folks on the shore, that allows us to learn anything at all - even in our own culture - beyond what we already know. Further, it is our location 'on the ground' (e.g., studying everyday lives) that puts us in a position to see people not simply as passive reactors to and enactors of some 'system,' but as active agents and subjects in their own history (p. 143).

A critical ethnography, I propose, is able to explore these agentive practices within the context of hegemony as it is articulated discursively and materially, for as Martin-Barbero (1993) argues, it is in this way that we may truly learn of the relationship between ideology and experience.

As ethnographic authority may no longer be assumed but must be *established,* I note here that my participation in the knowledge-production of this project has been informed by my experiences over the past twelve years as a volunteer youth worker in various civic and religious capacities with young people whose racial and class backgrounds differed from my own. Three working assumptions in particular emerge from my background and position. First, I am aware that when I am with teens, I am perceived as an adult and in many cases a cohort with the parents of the teens I have interviewed. Due to my class and race position

as an educated white woman, I am perceived much like one of their teachers. Second, I am aware of the importance of peer groups in the formation of individual and collective identity among teens. Third, in my experience as well as in my readings of adolescent psychology, it is clear that adolescents tend to be self-oriented, desiring opportunities to be "in the spotlight" on the one hand while also vigilantly avoiding any possible public embarrasment. Not surprisingly, therefore, I found that teens often enjoyed the individual interviews with me, as they perceived the interaction as an opportunity for unthreatening self-expression with an appreciative audience. I have attempted to account for these factors in my methodology, and thus I will devote the following section to describing how my research design evolved and how I believe it has enabled me to gain insights into the research problem.

The Research Design

As a result of grappling with these issues of knowledge, realism, and agency, I have chosen to employ a multi-layered research strategy which incorporates different audiences for and producers of identity narratives and differing levels of interrogation at individual, family, and peer levels of conversation. While I rely upon my own interpretations of the knowledge gathered as seen within a neo-Marxist framework, I am attempting through this multiply-situated research to gain as much insight as possible into how individuals present themselves when they are alone with the researcher, in the context of their families, or in the social setting of their peers. I am proposing that I learned of how these meaning-making narratives related to individual and familial

constructions of identity by exploring the ways in which people situate themselves within various discourses, examining in particular the narratives describing their relation to the public discourse of media representations concerning religion. The data-gathering and analysis has not happened in a linear fashion but rather has been iterative, encouraging a systematic process whereby I return to the field repeatedly not simply to build a "thick description" of a cultural practice, but to continually revise and refine understandings of how and in what contexts identity narratives are constructed.[5]

Rather than applying *sampling* logic to locating my research participants, I used a *replication* logic, as Yin (1994) recommends. Yin argues that "every case should serve a specific purpose within the overall scope of inquiry...Each case must be carefully selected so that it either (a) predicts similar results (a *literal replication*) or (b) produces contrasting results but for predictable reasons (a *theoretical replication)* (pp. 45-46)." Rather than representativeness or typicality, therefore, I looked for what Patton (1990) calls "information-rich cases," those from which "one can learn a great deal about matters of importance (p. 181)." Each of these are also *critical* cases in that they challenge expections of what a "typical" informant of a certain demographic group might embrace in terms of articulations of identity and media meaning-making. My project has adopted three approaches to the description and analysis of information rich cases: five case studies which have afforded an exploration of both the teen's family and peer contexts for describing media meaning-making practices, three peer-led discussion groups led by three teens from the five case studies, and three parent discussion groups which focused on issues

[5] "Thick description" is a term popularized by anthropologist Clifford Geertz (1973).

of intentions and practices surrounding media in the family context. I will describe each of these in more detail below. It is worth noting that in qualitative studies, validity is usually claimed in relation to the ability of the method to gather "rich, believable data." I believe that this multi-tiered approach, which began with individual interviews and then extended to groups of interviews, has to some extent allowed for the observation of an echoing of similar discursive themes across various times and contexts. While I believe that more and wider research could assist in confirming the modest findings outlined here, I believe that these observed overlapping similarities of discursive strategies do offer some measure of reliability to my study.

I would like to comment first on my inclusion of the role of religion in my sampling and analysis. As I have already noted, I believe that religion has constituted an important "blind spot" in the analyses of media in the everyday lives of people. Religion, although often equated with the institutions which have traditionally "housed" it, extends beyond these institutions, as discussed earlier.[6] Thus, this study does not wish to bracket out religion as something that is *a priori* of less importance in peoples' lives. Rather, it expects that there would be some relation between a person's religious world views and the framework of "meaning-making" he or she brings to the media experience. Moreover, while a focus on religious beliefs and practices in relation to media meaning-making might not seem the most obvious way to reveal the workings of hegemony within popular culture and its reception, I argue that it does contribute something unique to the current critical/cultural struggle in this area. Instead of considering religion as a subset within the

[6] Roof (1993) points out that a majority of Americans still report that they are "spiritual."

larger and hegemonic discourses of U.S. culture, it may be understood as a belief system that provides meaning to the hegemonic system as a whole to individuals and subsets of the culture. Moreover, U.S. religious pluralism has resulted in a multifaceted discourse of religion, as I will describe more fully in a subsequent chapter. When religion is expressed in popular culture, therefore, it is not simply "resisted" or "negotiated" by audiences, but is *regenerated* as it is incorporated in partial ways into the existing habits and practices of everyday life, thus giving shape and meaning to the larger hegemonic system itself for many of the culture's members.

My pool of research participants, therefore, includes persons from various positions *vis a vis* religious institutions, from young people who are more involved than their parents, to families who have drastically changed their religious affiliations and involvement, to families that have never had a religious commitment at all. In this way, I hope that my research may begin to shed light on how young people view their experience of religion, both within the context of their families and religious organizations and within the broader, public and mediated framework of a culture - as Martin Marty (1996) has termed it - "saturated" with religion.

I would also like to note the difficulties in using class as an autonomous category of analysis, as has often been the case in British ethnographic media research. Using Bourdieu's (1977) concept of the *habitus*, we may argue that an individual's location within culture - including past perceptions as well as one's immediate contacts with the family, neighborhood, or peers - informs meaning-making. Yet the

habitus which ultimately informs perception transcends all of these categories and is overdetermined by class belonging, Bourdieu argues. Such distinctions are more difficult to make in the U.S. than in the European context, however, for here they are largely held in place by cultural practices and prejudices rather than obvious economic or political constraints. Indeed, as Stacey (1991) pointed out in her study of working class women in the Silicon Valley of California, it is inappropriate to impose unexamined assumptions of class categories onto the lives of real people. Demott (1990) has chronicled the tendency for all Americans to see themselves as "middle class," which demonstrates the problem in analyzing ethnographic data based upon self-reports without the aid of macro social analysis. There is also the practical difficulty of categorizing individuals whose class identifications are less than clear. Consider the case of Elizabeth, a caucasian teen described more fully in chapter five, for example. While her household income and parents' educational backgrounds suggest a working class designation, the geographical identification with her neighborhood and school, her friendships with peers, and her divorced mother's current schooling which should raise both household and educational levels, suggest a middle class designation. In order to discuss class in this project and specifically in cases such as this, I adopted the pragmatic taxonomy articulated by Press in her study of women, class, and television viewing. Press (1991) argues that class is best determined by the employment of the father or husband. The mother's occupation is also taken into account in the case of divorced or widowed mothers. In cases of borderline employment, such as unskilled clerical work, education may be taken into consideration. In the case of Elizabeth,

then, I "split the difference," giving the designation of "lower middle class." This case illustrates, however, that a reluctance concerning class designation does not arise from knowing the people interviewed and seeing certain categories as derogatory, but rather in coming to know the intricacies of specific circumstances of the domestic contexts in the post-industrial age.

While the analysis I focus on in this project is based on the five case studies and six discussion groups I noted earlier, my study actually began with the interviewing of 70 individuals - 35 of which were teens - in 15 family groups around general issues of media practices, religion, and identity. All of the families in the study were recruited through contacts with various civic, educational, religious, and therapeutic groups as part of a larger study on media, the family and the life course, led by Stewart Hoover and funded through a grant from the Lilly Endowment.[7] Two or more separate interviews of approximately 90 minutes were held with each household: the first with the entire family, the second with individual members of the family interviewed separately and confidentially. I then explored the findings from these family groups in what eventually became five in-depth case studies of teens in the second stage of my research. In the third phase of my research, three adults and three teens were selected from the family groups to serve as leader/organizers in the peer-led discussion groups which focused on media use, the "readings" of popular media texts, and issues of religious pluralism, multiculturalism, and generational difference.

[7] My project, focusing on identity-construction among teens, contributes to this larger grant-supported effort. An additional six families were interviewed by an associate in that project, employing the interview guides I had constructed. My analysis concentrates primarily on the fifteen families with which I was directly involved.

The fifteen families participating in the research range in annual household income from below U.S. $20,000 (poverty level) to over $80,000 (upper middle class), and live in differing neighborhoods of the single metropolitan area of Denver, Colorado, an area of technology-driven economic growth and population influx, while also an important intersection of the more conservative and libertarian values of the Midwest and Southwest, respectively. Seven of the families are single-parent households, and one is a "chosen" family of a mother, her daughter, her daughter's two young children, her daughter's boyfriend and father of the children, her daughter's best girlfriend and her husband, and another male friend of her daughter's. I have defined family broadly as those sharing a household a majority of the time. The religious backgrounds of the families interviewed are quite diverse, as well. Five of the family groups claim some affiliation with a Protestant religion (ranging from Evangelical Protestant to more traditional mainline faiths), three families are Roman Catholic, two are Jewish (Reformed), one is Muslim (orthodox/Sunni), and four claim no religious affiliation. In four of the families, teens attend services at religious organizations which differ from those in which their parent(s) are involved.

At this point, I would like to reflect a bit on how the methods of the second and third stages came to be employed. This will then lead to a more extended discussion of the demographics in both the case studies and discussion groups that formed the bulk of my research.

Evolution of the Case Study Method: A Failed Research Effort

Because I am committed to a self-reflexive research project, I feel

that it is important to begin with a recognition of the fact that my research design was not fully "operationalized" so as to guide my knowledge-gathering from the beginning, but rather the process of conducting interviews, presenting initial analyses, learning of similar research projects, and interacting with various people (both interviewees and scholars) informed my own understandings of what I hoped to learn, and thus helped to shape the research design over the course of a two-year period. In fact, I was not even committed to the importance of a self-reflexive stance when I began this project, although the realization of the possibilities in such an approach evolved quickly as my research unfolded.

As part of the Lilly project, Stewart Hoover and I were given an opportunity to meet with various scholars on how best to approach the topic of interest to that study's: the changing nature of religion in the media age, and the embedded role of media in the meaning-making practices of family life. One of the persons with whom we met was Judith Stacey. I had been impressed with the depth-research approach taken in her text *Brave New Families* (1991), and after meeting with her was convinced that by getting to know a few teens quite well, my own perspective on the questions of meaning-making and identity-construction could be greatly expanded, therefore contributing to a better analysis of the issues. I decided to call one of the families that had been most intriguing to me during my initial interviews with them: Amber and Joe, the caucasian single parent and his 16-year-old daughter introduced in the previous chapter. They were quite willing to be interviewed again and to become involved in a more active research relationship. Over the course of the next few months, therefore, I was in

fairly regular contact with both of them, listening to music, watching television, and mostly providing an attentive audience to Amber's dramatic life stories. I even had an opportunity to meet with Amber's extended circle of family and friends when I attended her 16th birthday party. While my relationship with Amber was developing into one of trust and possibly even mentoring, I had increasing concerns about a potential falling-out with Joe should he learn that I had known of Amber's involvement in a gang long before he had. I was completely unprepared, however, for the entrance of his ex-wife into my research agenda. Here are comments from my field notes:

> July 27: Joe called back Wednesday morning (7/24)...he said, sounding embarrassed, that he had another thing he wanted to talk with me about (we had been discussing Amber's birthday). He said he and his ex-wife Leanne had gotten to talking about this interviewing that I was doing and she got all bent out of shape and said she wanted to make sure I heard *her* side of the story. So, Joe said sheepishly, would I mind interviewing her sometime for my study? Would I mind?! Since I'd been trying to figure out how (and whether) I could ask to do this, you can imagine how much I had to conceal my happiness at this serendipidous offer. I said, yes, I could interview her, and I reminded him that it would have to be confidential and I wouldn't be able to tell him what she said, nor would I be able to tell her what he said. He said sure, he knew that. "She's not a very nice person," he said, and then gave me her number and said I could "form my own opinion." I reassured him that I couldn't be totally impartial since I already had a good relationship with Joe and Amber and knew about their troubles with her (I wanted to be clear that my loyalties were with them). Well, he said again, "she's not a very nice person."

When I called Leanne a few days later, I learned that she had been quite upset that Joe had agreed to Amber's participation in my interviewing process without consulting her first. She immediately demanded copies of the transcripts of the interviews I had done with Joe and Amber. I explained to her that I couldn't give her that because they had agreed to be

interviewed under the knowledge that I would keep their responses confidential, but I could send her a copy of the questions I asked, if she wanted that. She became impatient and told me that she and Joe share the guardianship of Amber, which meant that her permission was required for "anything that involved Amber - from things like this, to dating, to everything." I told her that I had understood Joe had legal custody of Amber, as that is how he had explained the arrangement to me. I then explained that my study was based on families that lived together for the majority of the time, as Joe and Amber did, and that I would be glad to interview her family in the same way and with the same confidentiality, but because she had not been present at the initial interviews and did not share their household, I could not give her the transcripts. I reiterated that I could send her information on the study along with the questions I asked, and she asked that I send those to her right away. She then mentioned that her sister had just finished her Master's thesis, and she was therefore familiar with the amount of effort required in my own work. She said she knew that I had been unaware of the legal need for her permission to interview Amber, and that she "did not want to cause trouble" in my study. Then she gave me the address where I could send the questions and information on the study. Although I had heard the legal threats implied in her comments, I reiterated that I would be glad to hear "her side of the story," as I knew that there were always many different ways of looking at family relations. She said she thought that "wouldn't be a good thing to do," because she and Joe look at everything so differently. She again thanked me for "getting back to her so quickly," because she had "only found out about this earlier this week, and that was

by accident."

After I hung up, shaken, I suddenly felt like I was in over my head. I recognized that my research project had become one more pawn in the ongoing battle between Joe and Leanne. Reconstructing the events, I would guess that perhaps Joe had bragged to Leanne that he was being interviewed to emphasize his competency at parenting (something she had insisted he could not do). She in turn saw her exclusion from the interview process as an opportunity for Joe to denigrate her once again to other "public" figures (he had been featured in a newspaper article about single parenting months earlier). Her interception into the research project allowed her to reassert her authority as Amber's parent. Based upon her threatening tone with me and their history of escalated conflicts, I decided to call my advisor, Stewart Hoover, to discuss the handling of the situation. As he was worried about the possibility that Leanne might sue the university to obtain the transcripts, he strongly advised me to cease my relationship with Joe and Amber. Even a formal cessation was risky, as any way in which I might attempt to tell Joe that the research was to stop might further escalate the conflict and put the larger research project at risk, which would in turn possibly jeopardize its funding.

I went to see Joe and Amber one last time on the day of that fateful phone conversation, a visit which had been prearranged. Because Joe did not ask if I had had an opportunity to call his ex-wife, I did not mention it. I left that evening with a vague mention that I'd see them again. My last contact with them was a brief postcard I sent to them while on a trip a few weeks later. I never followed up to schedule the family interview with Leanne.

Six weeks after that fateful phone call, I still was not reconciled to the abrupt ending of the research relationship. I wrote in my field notes:

> 9/6: I still feel angry with Joe's ex-wife Leanne...Of course it's hard for me not to "side" with Joe and Amber, but it's not like I haven't dealt personally with being diplomatic among feuding spouses! (my parents were permanently separated under contentious circumstances less than two years before this study began)... The interesting thing is that probably I have more in common with Leanne in her comfortable middle class life than I do with Joe and Amber (who struggle financially). I suppose one learning from this might be that class, while narrowing options for social interactions, isn't totally determinative of who you end up liking, and why. I liked Joe and Amber, even while acknowledging that in some ways I didn't really understand them, just as I always felt with my mom's side of the family... I guess it's interesting, in a way, that Leanne should be less trustful of an academic like me than Joe and Amber seemed to be. ...Still, I wish I could talk with her personally. ...

Now, reflecting on the experience more than a year later, I feel no desire to communicate with Leanne and have in fact achieved some "critical distance" from the research relationship, although in principle I can still see the value of gaining insight into the two different households between which one teen was shuttled. However, I do wish that I could be in contact with Joe and Amber, as I feel confident that the developing research would have been much richer with their inclusion. Among other things, the interaction with Leanne had caused me to reflect more deeply on class differences, which I felt could have been helpfully explored with them, as well.

It took several weeks (and an impending presentation deadline) before I could recognize the positive aspects of the discontinued research. On the one hand, while I felt that the work was cut short, enough interesting data had been generated to warrant its examination as a case study. On the other hand, the reseach had been discontinued fairly early

in the overall data-gathering process for this project, which allowed me to locate a second family as a case study.

The Second Case Study: A Case of Serendipity and Surprise

During my initial interviews with the Farleys, a single parent lower middle class caucasian family, I had not even considered that my relationship with them might continue beyond the initial two interviews. My hesitation grew mainly from the skepticism Jean, the mother, had seemed to demonstrate toward both me and the research project. However, 14-year-old Elizabeth struck me as capable of reflecting upon and articulating her teen experience at a more sophisticated level than any of the teens I had interviewed thus far. At the end of my second interview with her, therefore, I had asked her if she would be interested in serving as a "consultant on teen life" for me. Even I wasn't exactly sure what that would mean at the time, but I felt compelled to leave the door open for later contacts with her. I had been looking for a teen who might be capable of participating in a methodological experiment which had been suggested to me by Elizabeth Bird (1995), and which eventually became the third phase of my research. Addressing my concern about the difficulties of establishing trusting and revealing research relationships with teens, Bird had suggested that perhaps I should recruit and train a teen to serve as a leader of a focus group on media practices. Because the Lilly project was interested in mediated practices of home life, I proposed that perhaps Elizabeth could organize and lead such a group on a discussion of telephone use. Elizabeth agreed to meet with me again to discuss this possibility, and over a period of the next six weeks, I met and talked with

her by phone to make these arrangements. After much discussion and training, Elizabeth held the group without me present and then met with me a few weeks afterwards to give me the audiotapes for transcribing. Serendipidously, my relationship with Joe and Amber was to end within weeks of this group's meeting. It was not until I listened to the audiotapes a few weeks later that I realized I had already begun a second case study, almost without noticing:

> 9/6: This is how I become aware of my methods, and what I'm learning: it suddenly occurs to me that I have been in contact with Elizabeth several times throughout the summer as I've recruited and then trained her to serve as both a discussion group leader and a transcript reader. I guess I have a somewhat more extended relationship with her already established, and should probably parlay this into a more formal (at least in my mind) long-term research relationship. Du-uhh.

In my field notes, I acknowledged that three things contributed to the fact that I had not recognized the potential of my relationship with Elizabeth. First, I had approached our relationship rather instrumentally: I was, after all, paying her to lead the peer-led discussion groups, and I was viewing her primarily through her task as "data-gatherer" for me. Moreover, she'd taken the task quite seriously and I often felt that she spoke to me as an employer. Thus our relationship had been defined less by my interest in her own experiences, *per se*, than by my interest in the work she was doing for me to help me gain access to teen group situations and discussions. Second, I was already busy trying to meet with Amber and Joe on a regular basis, and Amber, being two years older than Elizabeth at the time, was (not surprisingly) more mature. She was also clearly interested in promoting a relationship between herself and me, as she appreciated the interest I had taken in her. Grieving over the abrupt ending of my relationship with Amber and Joe was what actually brought

my attention to the potential of my relationship with Elizabeth. Third, as noted above, I had been concerned about Elizabeth's mother and her skepticism toward me. Yet as my relationship with Elizabeth evolved, Jean came to see me increasingly as a trusted friend of her daughter's. This had the added and unexpected benefit of placing my family role in a very different location than with the other families I was interviewing, as it was always the parents who were my primary contacts there.

After I realized that I had already inadvertently established a relationship with the Farleys, I needed to redefine myself from employer to researcher and friend. Elizabeth was receptive to continuing to meet with me, and over the course of the next two years we continued to see each other, sometimes informally for social occasions, and at other times for more formal interviews. We also communicated frequently by phone and on e-mail, which, as will be clear in the chapter focusing on Elizabeth as a case study, became a prominent part of my learnings from Elizabeth. A year after the initial experimental peer-led discussion group conducted for the larger study under the direction of Stewart Hoover. Elizabeth and I worked on another discussion group to address the questions of religious identity, media use and tolerance: the first peer-led discussion group of the three I eventually held with teens for the current project. My relationship with Elizabeth, while occasionally still formally related to my research, has evolved into a mentorship: we now talk much more frequently about her homework assignments and other school challenges than about her media use, although of course the latter still emerges in conversation quite often.

The third case study: Michael

As was true in the previous two case studies, it was after the initial two family and individual interviews with the Pearsons that my relationship with 15-year-old Michael emerged as a possible case study. Yet unlike those families in which I had met the parents first, Michael and I met through a friend of mine, a television producer who had "discovered" him as a young and articulate African American with amazing competence in web design. Michael had been featured on an educational television program on innovative community computer projects, and we met at the program's "premiere," which was held at the Governor's mansion. After a conversation about his interests, Michael agreed that I could formally interview both him and his family. It was during that interview that I discovered that Michael's mother credited her religion with keeping Michael from the gangs and other negative distractions that claimed so many of his colleagues in his underprivileged neighborhood. Michael himself reported to me that his favorite television program was *Touched by an Angel.* I surmised, therefore, that he might be more capable of leading a group discussion about this program than other male teens his age, and I asked him if he would like to do this. He agreed, and I approached the training and follow-up for the peer-led discussion group much more intentionally than I had with Elizabeth as an opportunity to establish a more long-term relationship.

My research relationship with Michael has raised perhaps the thorniest issues I have yet encountered regarding the researcher's ethical responsibilities as a result of the relationship. As should be clear from my earlier critiques of the humanistic focus in early U.S. media ethnographic

work, while I am supportive of the research goal of "giving voice" to persons in society who are usually silenced, I am also skeptical that such an approach might have an observable or direct benefit with my research participants themselves. My experience has been that any benefit arising from the interviewing situation is largely therapeutic; after all, I am not enabling them to have greater financial, political, or social control over their situations.

Moreover, Michael is a bright and calculating individual, and it is clear to me that he is aware that while I may not be in a position to directly offer or find him a job, I am at least positioned to know of a greater possibility for contacts in the entertainment industry than he himself would. In feeling some sense of responsibility toward helping him to achieve his "dream" of becoming an actor, I am aware of the possible imperialistic tones of the situation, as numerous critics of ethnography have pointed out (See, e.g., England, 1994). Yet I feel that it would be dishonest to pretend that I am not so positioned to help, and I believe it would in fact be unethical to not utilize my contacts and connections when possible. I have thus made some contacts with a talent agency on his behalf, and continue to promote his name whenever I learn of possible jobs in the entertainment field. In the meantime, we continue to meet together informally when his busy schedule permits.

The fourth case study and another failed effort: Jake

I had now begun to explore two case studies, both of whom were articulate teens with differing life experiences who would identify themselves as "religious." I felt, therefore, that I really needed to gain

some insights from a teen who might *not* self-identify as "religious": what might be his relationship to the public discourses of religion? When I interviewed Jake while pondering these questions, I thought that perhaps he could provide me with some insights into the question. Jake, a 17-year-old caucasian male from an upper middle class blended family, was literally from the "other side of town" from Elizabeth, Amber and Michael. Moreover, he had never attended the services of organized religion, although his mother had been active in a conservative Protestant congregation until her divorce from Jake's father when Jake was very young. Despite the lack of religious identification, however, the family frequently spoke of "guardian angels" and the vague idea that God existed because, as Jake says, "there's been a couple times where it's been close, where you think you're gonna die or something, but then somehow you get out of it."

Jake lives with his mother, stepfather, and stepbrother, seeing his own father rather infrequently. He is in regular contact with his stepfather's other two sons, however, as they all share an interest and weekly participation in dragcar racing. During the summer months, the entire family, occasionally including his college-age sister, heads to the track for Friday evening races. In addition to racing, Jake works at a fast-food outlet a few hours a week. He also spends some of his spare time on his artwork, as he and his mother share an interest in freehand drawing.

After my initial two interviews with him, I asked Jake if he would be interested in serving as a peer-led discussion group leader for me, as that had become the entry point for my case study research with Michael. He agreed, and we then met to discuss the logistics of the group and to

spend an evening watching television together. Yet months passed and he did not organize the group. After numerous prodding phone calls, organizational suggestions, and subtle pressure of deadlines, I finally provided a reasonable excuse for him to bow out of the responsibility, which he chose to do. We met again for a brief discussion, but his embarrassment about not getting the group together coupled with my own impending deadlines meant that while I left the relationship open for further contact, nothing has yet been pursued.

I should note here that in the cases of Amber, Elizabeth, and Michael, I had trusted an intuitive feeling that each could participate fruitfully in my research project and that my relationship with them could offer some reward, in turn. Yet with Jake, I have to confess that my interest was rooted first in his demographics as an upper middle class non-religious male.[8] The fact that he was less articulate than the others concerned me, but I believed that he was still reflective and was not uninterested in the questions I asked, and also had offered explanations that had contradicted my expectations often enough that I found him to be an "information-rich" and critical case. Further, I knew that his mother, who placed high expectations on him, would probably goad him into helping me with the research. I was concerned about the possibilities that a teen male might find inordinate social pressure against organizing a group of friends for ordered conversation, but was willing to experiment. Yet Jake confirmed my concerns during our last meeting, agreeing sheepishly that he had been reluctant to organize a group, as the content

[8] The fact that he was caucasian was not one of the factors in selecting him, although his family's sentimental religion might be related to his racial/ethnic and class background. Further research would need to bear this out, although Roof's (1993) findings are suggestive of this.

and the organization itself could of course be conceived of as incredibly "uncool" among his cohort.

The fifth case study: Hasan

As soon as I started to suspect that Jake might not come through for me as a peer-led discussion group leader, I began to think about calling Hasan, yet I hesitated. I had already established a research interest in his biracial orthodox Muslim family, as I had conducted both in-depth interviews and a participant observation with him and his siblings the year before. I understood that they (particularly Hasan's mother) had seen it as worthwhile and in fact consistent with their religious commitments to tolerance to educate me on their practices, so I knew that my request for further assistance would probably be well-received. Yet I anticipated some discomfort: as I had already employed the television program *Touched by an Angel* with three previous teens and two groups, I was quite interested in continuing to use the show, yet wondered if Hasan would be offended by what I believed was its overt Christian message. Thus, when I finally called Hasan, I explained my intentions and concerns over the phone, but he agreed enthusiastically to meet with me nonetheless.

It had been about a year since my previous contact with Hasan and his family, and in that year each of the four Ahmed siblings had matured considerably. Hasan himself, now at 15, was much more articulate and self-assured, and like his mother, saw himself as educating me on the differences between his own beliefs and those he saw represented in the *Touched by an Angel* episode we watched together. In fact, by the end of our first meeting I was convinced that having someone like Hasan, who

clearly saw himself constructed as an "Other" in relation to the popular media text I'd chosen, would greatly add to my research design. Hasan, sensing my interest in this very issue, intentionally gathered a racially diverse group of friends for the peer-led discussion group. The friends had differing levels of commitment to organized religion, as well.

Because of the family's religious beliefs which encompass some concerns regarding gender issues, I was quite conscious of how I might construct my ongoing relationship with Hasan. Thus, most of my meetings with him were in the central room of the house, where his mother and siblings were nearby and often entered the room during our conversations. At least one of his siblings joined us when we went out for meals together, as well. I admit, therefore, that I was much more interested in respecting the family's beliefs than in maintaining a strict rule of "confidentiality" between researcher and research participant. The fact that Hasan established the conditions of the interview in this sense only underscored my growing sense that the ethnographic process is *dialogic,* and that participants do in fact have some power in the relationship as they may control their contribution to it. Of course, I still accept the fact that the ethnographer maintains the power to edit, interpret, and construct the research project in the first place.

Although I have mentioned the peer-led discussion groups and their role within the case studies, I would like to make an argument for why I employed this method and how this contributes to the overall project.

Peer-Led Discussion Groups

As Lunt and Livingstone (1996) have noted, focus group methods in media audience research have been used for both critical and institutional purposes beginning with Merton's and his colleague's (1956) employment of the method as a supplement to quantitative approaches. Much like the ethnographic media research discussed above, focus group research has assumed an apprehendable but largely inaccessible "reality" beyond the interview situation, and thus focus groups are designed for "getting in tune with the reality of the interviewee," as one textbook on focus groups notes.[9] Moreover, like quantitative researchers, some focus group organizers strive for the "objective" distance between themselves and their interviewees, noting that pre-formed social relations may interrupt this process. As one textbook notes:

> Caution should still be used when considering focus groups with close friends, family members or relatives, or work groups. People who regularly interact, either socially or at work, present special difficulties for the focus group discussion because they may be responding more on past experiences, events, or discussions than on the immediate topic of concern (Krueger, 1994, p. 7).

It should be clear at this point that my interest is in coming to a greater understanding of just these contexts of past experiences and other relevant factors and the ways in which they mediate meaning-making practices in the group's discussion, for I believe that these inherently shape the narratives that emerge, whether in the group or beyond it. My own design, therefore, includes groups of people who knew each other prior to the group's gathering, follows the work of Liebes and Katz (1990), who employed this approach to simulate social occasions in which similar discussions about television programs might emerge. Liebes and Katz, as

well as Radway (1987) in her work on romance readers, were interested in approaching participants *as* members of a social group, rather than as *individual* representatives of certain subcultural groups. Additionally, I believed that as group members might be expected to agree on many of the topics discussed, they would also be more likely to voice dissenting opinions and engage in active defense of their positions due to the already-established trusting nature of their relationships.

While I discussed the differences between critical and "naturalistic" approaches to ethnography earlier, there is still one remaining troubling aspect of naturalistic inquiry which has not been mentioned up until this point. Too often, these approaches describe findings in terms of latent social and psychological theories and assumptions, which limit their analyses to psychological, and too often individualistic, variables. An alternative approach, advocated by Jensen, is to explore the role of media in peoples' lives by analyzing how they *talk* about media (rather than solely analyzing how we assume the media cause them to *behave*). As he argues,

> if mass communication research is to contribute a critique of media in the interest of their audiences, it must develop methodologies that engage audiences in a process of reflexivity about the social purpose of mass communication (Jensen, 1995, p. 95).

While Jensen meets this challenge through workshops on the future of television programming, my research takes the same approach in a different direction. Instead of exploring the behavioral effects of media on teens as is common in the literature, I encouraged teens to reflect on their media practices themselves, thus allowing them to engage in a level of reflexivity that they often have not previously encountered.

Furthermore, by working with individual teens to jointly develop a question guide for the peer-led discussion groups, I allowed teens to both engage in reflection on their own practices as well as on how they interact with their peers about these practices. This reflexivity is what made the training and interpretation of the group's discussion such a vital part of the case studies I developed. The interaction with the leaders, as well as the group's conversations, then expanded the individual framework to a social dimension, thereby further exploring participation in language as a social act.

The fact that the teens led the groups without any adults present is also significant, of course. As noted earlier, when we consider the dialogic aspect of both ethnographic research and identity-construction more generally, the role of the audience for such constructions becomes crucial. By having the teens engage in the group without adults, I did not simply want to attempt to replicate an "actual" conversation among teens. Rather, I wanted to turn over some of the authority of ethnographic inquiry to the teens, and to my peer-group discussion leaders in particular. As I believe that teens are most comfortable conversing when they can control the terms of the conversation, I wanted to facilitate this control. The results from the group, therefore, were sometimes disappointing to me, as I did not learn exactly what I had anticipated. Upon first listening to the tapes, for example, I felt disappointed with interruptions such as "She already answered that. Next question," or outbursts of uncontrollable laughter, or even the failure to ask what I believed were "obvious" follow-up questions. Yet I grew to sense that the content of the conversations were guided to a much greater extent by the leaders' (and

group members') varied levels of comfort with and interest in the topics than would have been the case had I been present to moderate. This encouraged me to pay closer attention to the nuances that were present in the conversations, such as the various cues that caused embarrassment, a quick change of topic, or loud and repeated efforts at being heard by one or more participants.

I should note, too, that each of the peer leaders handed over the tapes with some embarrassment, concerned that I would not be able to "get much out of it" because their group members "didn't take it very seriously." Yet I was surprised that this embarrassment was not present at all while we listened to the tapes and read through the transcripts together. I think this is evidence of the fact that the teens sensed the somewhat ambiguous nature of the research endeavor. They correctly surmised that as an adult researcher I would have some specific desires for coherence, clarity, and seriousness, yet they also sensed that the group's talk did, in fact, represent "how teens talk about these things," which is exactly what I had expressed an interest in.

One of the unforeseen benefits of these peer-led discussion groups was that while they were homogeneous in many ways, they mirrored social groups that do not necessarily share the same religious commitments. Elizabeth's group, for example, consisted of two very active church members (Lutheran/ELCA and Presbyterian [USA]), two girls with marginal church affiliation and attendance (one Roman Catholic and one who attends "various churches"), and two who had never attended religious services. These differences allowed for negotiation over religious meaning, as I will detail in a subsequent chapter. Yet there were,

of course, similarities among the group. Each of the six girls was caucasian and attended the same school, and only two were not active in any school organizations. All but Elizabeth were in two-parent families that would be considered middle class due to parental occupations and educational backgrounds.

In Michael's group, three of the boys attended religious services regularly although they were not actively involved in their congregations (two Jehovah's Witnesses, one Church of Christ), and three never attended religious services at all. Each of these boys were African American and lived in the same lower-income neighborhood, although they did not attend the same school. Three lived in a two-parent household, two were in single-parent (female headed) households, and the last lived with his grandmother and three uncles. As was the case with Elizabeth's group, the varied religious backgrounds provided an interesting grounds for analysis, particularly as much of the religious/inspirational talk in this group was agreed upon consensually.

In Hasan's group, two teens were actively involved in their religious instititutions (Nazarene and the Colorado Muslim Society), two more regularly attended religious services but were less involved (Evangelical, two Catholics), and one did not attend religious services at all. As noted earlier, his group was racially diverse, including two African Americans, two biracial African American/caucasians, a Korean/Philippino, and Hasan himself is a biracial Arab African/caucasian. Each of the boys attended the same school and could be considered lower middle class when considering parental occupations and educational levels. Three live in two-parent families, two in single parent

(female headed) households, and one lives with an aunt, uncle, and grandmother. As this group was the least homogeneous, it also offered the greatest challenges for interpretation, although several patterns in the group's discussion of issues of religion and values were observable.

As noted earlier, the peer discussion group leaders were paid nominally to work with me to develop an appropriate group interview guide, recruit six friends for the group, receive training from me to lead the group, and finally assist in the interpretation of the transcript of the group's discussion. The parent's groups were conducted quite differently, for I believed that their assumptions governing research might make them *more* comfortable with an adult researcher present to guide the conversation. Moreover, it was easier and more comfortable for the adults to conceive of themselves as *consultants* to my research rather than as "objects" of it, an approach more consistent with my dialogical notion of research and knowledge-construction. Thus the procedure for the organization of the parents' groups was much different than the teens' groups had been. For one thing, I waited until my work in the interpretation processes of the teen groups was well under way, so that I could ask questions that were pertinent to my findings and consistent with my emerging questions. Secondly, I selected three families with teens who I believed represented differing backgrounds and interests and then asked if they would organize a group of friends at their home for the discussion, with the understanding that I would lead the group. I felt that training parents to serve as group leaders would have introduced several obstacles, not the least of which was a willingness and ability to devote the time needed to participate in group-leading training. Moreover, I feared that

even a well-intentioned parent might "steer" the group discussion in a particular way due to interests that may or may not have coincided with mine. I finally concluded that participation in the research dialogue offered more benefits among parents than it had among the teens.

The first family I called were the Ortizs, an articulate lower middle class Mexican-American couple with conservative Roman Catholic values who live in a border area between the Denver's large Mexican American community and a caucasian working class neighborhood. The Ortizs had expressed a willingness to participate further in my research when I'd interviewed them individually and jointly nearly a year before, and when I called them, Judy Ortiz enthusiastically agreed. I believed that the fact that I offered to pay the organizing couple $75 and each participant $25 was a motivating factor, although they were also quite interested in the research project, as well. Theirs was the first group to be held, and the group consisted of Judy and Tom Ortiz, Judy's recently divorced sister, another divorced sister who lives with her boyfriend, and a couple who, although separated, are raising their teen daughter together. Each of the parents were Mexican American, and Judy's sister's boyfriend was African American. Although Judy and Tom have what might be considered a middle class income, none of the group members had completed college and the separated couple individually earn close to poverty wages.

The second parental group I'd called was the Kuczynskis, a liberal caucasian couple who were members of a United Methodicst church and lived in a neighborhood known for its progressive views. They eventually decided to cancel, however, as they were unable to generate enough interest among their friends to pull a group together due to busy

schedules. Most of the fifteen family groups I had interviewed were fairly conservative in their religiosity, and I had wanted to interview a group of parents with liberal political and social commitments so as to contrast with the Ortiz group. This led me to invite Lester Gray, a friend of mine and who is a former Presbyterian pastor and middle class caucasian parent of two teenagers, to assist me in putting together a liberal-minded group. Lester had expressed an interest in the research project during informal conversations we'd had throughout my work. On the morning of the scheduled group meeting, however, Lester called to let me know that we would need to reschedule as a number of his participants had backed out at the last minute. I asked him if he thought that perhaps I should have offered a higher payment for incentive, but he believed that the amount of money had not been a factor, but rather that people had cancelled for extenuating circumstances.

After a few more reschedulings, Lester and his wife Ellen hosted the second group in their home on a Sunday evening. In addition to the Grays, the seven-member group consisted of two other married couples and a woman whose husband was not able to attend. All of the group members were caucasian with annual household incomes above $70,000 and at least one spouse working in a professional/executive occupation. They all live in the same caucasian suburban area of the city, and none had ever been divorced or separated. They knew one another as former members of a Presbyterian (U.S.A.) church that had experienced significant turmoil the previous year, and thus while they all self-identified as religious or spiritual, none are currently members of a congregation.

The third person I contacted to organize a group was Dottie

Donahue, a caucasian single parent in a poor family who had expressed what might best be termed 'alternative' religious and social views during my interviews with her and her family members. While the offer of payment may have been a motivating factor, it seemed clear that Dottie was willing to organize the group primarily as a personal favor to me.

For various reasons, Dottie had a difficult time getting a group of friends to commit to participation in the group study. After weeks of rescheduling, I arrived at her house at the appointed time, but no one else had arrived after a half-hour of waiting. Dottie and her daughter Deedee then began calling the various people they'd asked to participate. One had forgotten about a previously-scheduled appointment, a few did not return the phone calls, and one opted to watch a television program at home instead. After an hour and a half, Dottie's sister arrived. At that point I asked Dottie whether or not we should reschedule, but she suggested we wait to see if anyone else would arrive. After another half- hour, I again suggested that we reschedule, and Dottie agreed.

This failed group presented several methodological challenges for me. First, I was under a tight timeline and was anxious to get the groups under way. Thus while I was patient during the scheduled meeting time, I also wanted to attempt a rescheduling as quickly as possible. Second, and more importantly, I appreciated Dottie's good will in offering to help me organize a group, and thus I wanted to be careful to maintain our good relationship. I am aware of the differences in our background, however, and perhaps it is because of this that I was rather unsure of how to respond to the lack of the attendance of the group. As I wrote in my journal that evening:

2/16/98

I wasn't at all sure how to respond to this lack of attendance. One way: be authoritative, or manipulative. Subtly direct her in how she should organize the group along the lines of my own middle-class expectations. Yet then I thought: how do I know that her friends would have come if they'd had more time to plan, or if she'd called to confirm (because Dottie and Deedee did do this from about 7:30 to 8:00), or if she'd been more definitive about it? Another way: express disappointment. But I was afraid that then I'd been seen in sort of a teacher role, someone associated with education, judging her for not "measuring up." I didn't want to do that, and didn't think I could express either anger or disappointment without having her think they were directed at her or her friends. So I sounded kind of nonchalant, forgiving, oh well, agreeing with her suggestion that we'll try again next week.

Unfortunately, a family emergency forced me to cancel the scheduled group discussion a week later, and thus I do not know whether or not it would have materialized at that time. However, we attempted to reschedule several more times and finally I needed to move on with my research without this group's input. Because I am still interested in hearing from her group, I hope that perhaps it will be organized sometime in the future, for an upcoming research project.

The day after the failed group, I decided that it probably would be a good idea to organize a fourth group just in case Dottie's group did not materialize. I called Cathy Swenson, a caucasian lower-income single parent I'd interviewed nearly two years earlier when I was just beginning my research. Cathy enthusiastically agreed to organize a group, and again, although I sensed that she was interested in assisting in the research, the offer of payment to both her and the participants seemed to be a motivating factor.

All but one of the members of Cathy's seven-person group is involved in a support group for single parents. As I had made several

research contacts through this same organization, one of the group members was Jean Farley, Elizabeth's mother, and another was Joe Dearborn, Amber's father, who I had not seen since I had discontinued contact with them after the veiled threats from his ex-wife. I had also interviewed Sandy Lipton previously, but had not met Dick, Justine, or Donna before. This made the group dynamic a little different than the other groups, as they were quite interested in learning of my research findings and were more familiar with my research goals going into our time together.

The members of this group live scattered across several western and northern suburbs of the metro area. While two of the members have annual houshold incomes between $30,000 and 69,999, the other five reported incomes below $29,999. They all own their own homes (which in Joe's case is a trailer) and all have at least some college, but only one has a completed degree. All are caucasian except Joe, who is part Native American.

This group has differing religious affiliations: an observant Roman Catholic and an active member of a nondenominational church, an inactive member of a Lutheran church, an occasional attender of a Baptist church and a person who attends both United Methodist and Catholic churches occasionally, and two who have no religious affiliation but who self-identify as "spiritual" and "a believer" respectively, the latter a leader of Bible studies for his children.

The teen and parent groups followed the same basic interview structure, beginning with general demographic questions, watching an episode of *Touched by an Angel,* and then discussing first it and then

broader issues of interpretation and positionality. While the teens watched the entire episode of the program, I showed only the final ten climactic minutes with the parents, affording greater time for discussion while still giving them enough of the "text" to enable comments on it.

Data Analysis

The first *formal* stage of my analysis of the data involved the writing up of a case study for each person and/or family interviewed, creating a description of how each is positioned *vis a vis* the social groups mentioned before, and how each person's narrative illustrates this positionality in discussions of the media. I presented initial analyses in three forums and received helpful feedback in each (Clark, 1996a, 1996b, 1996c). The comments from professionals and colleagues enabled me to refine my analyses, and also assisted in planning for revision of the interview instruments. In the second stage, I sought patterns in the data that I believed might illuminate common themes which resonated with my research participants. I also explored similarities between my findings and those in other studies that asked similar questions. This is similar to the approach laid out by Jensen (1995):

> In reception interviews, it is especially rewarding to establish the central arguments of the respondent, their interrelations in the context of the whole interview, as well as the substantiations and implicit assumptions that are taken by the respondent to support the arguments. In linguistic terms, the three major analytical categories are coherence, presuppositions, and implicit premises (p. 78).

I am less convinced than Jensen of the "coherence" that emerges from ethnographic narratives, however, as I will discuss in more detail in chapters four and seven. While there are certainly themes that emerged

in conversations with the research participants, I found that people were often inconsistent, and sometimes even had elaborate justifications for these seemingly contradictory views. I found, in fact, that it was the instances of incoherence and inconsistency that often best illuminated the fissures in the larger hegemonic discourses in which they were engaged. Thus, as noted earlier, paying attention to the gaps, abrupt changes of subject, or moments of embarrassment were particularly fruitful moments in the data analysis.

The third stage involved the greatest abstraction from the data, as I attempted to match these patterns of talking about representation with what I believe are significant discursive strategies. This required that I pay as much attention to what was said as to what was *not* said. The analysis at this stage sought to develop theory as it strove to answer questions at the discursive, rather than simply conversational or interpersonal, level. This required a sensitivity to the ideological forces colluding to naturalize certain viewpoints which are expressed both in the media and among audiences. At this stage, which became the foundation of chapters seven and eight, I borrowed my models for analysis from textual and critical/cultural analysis. My approach at this latter stage was similar to that advocated by Jensen (1995). In his discussion of scientific analysis, he argues that there are five levels of such analysis:

> The discourses of mass media, and of audiences about the media, are the objects of analysis at the first level of everyday discourse, to be grasped and documented in the categories of analytical discourse, for example linguistic discourse analysis or content-analytical coding. The third discursive level specifies a methodological discourse in terms of research designs, analytical procedures, and bases of inference. Further, the findings about media and audiences are necessarily interpreted in the framework of a theoretical discourse. Finally, the status and explanatory value of the other discourses must be justified at the level

> of epistemological discourse. Each level makes up an interpretant in the chain of scientific semiosis (p. 62-63).

The primary differences between Jensen's proposed scheme and my own is in the approach to knowledge generated from the data-gathering process. Thus the order in which my research process took place differed from Jensen's. For instance, I felt that it was most helpful to begin by exploring and explaining epistemological, theoretical, and methodological assumptions which shaped the research design. After describing the everyday discourse, I categorized it analytically and then finally attempted to interpret it within what I understand to be the broader theoretical discourses of media and society. Thus while the descriptive everyday discourse informed my development of theory, I was also conscious of the ways in which epistemological, methodological, and theoretical assumptions informed the very process of my learning from the beginning. This awareness differentiates my own approach from the "grounded theory" approach which purportedly begins with empirical observation (Glaser & Strauss, 1967). Because I believe it is impossible to practice such observations without an implicit theory, I prefer to acknowledge those theories from the beginning.

I want to make it clear, however, that my research did not proceed neatly from one "stage" of analysis to another, but rather was a constantly evolving back-and-forth process which included conversations with research participants and scholars. Although it may have been "tidier" to do my theoretical research, collect all the data, and then finally analyze it in light of the theory, I found that this "messier" approach actually allowed my findings to evolve dialogically and, I hope, the review of this

process here highlights how I gradually gained the ethnographic authority necessary to tell a partial story that contributes to academic knowledge.

Chapter Four

Adolescents as Selves, Adolescents as Audiences

One of the reasons I chose to focus on adolescents in this study is that as they struggle to articulate abstract ideas that have only recently become of interest to them, they are often less aware of patterns of thought than their parents, and hence less inhibited by social sanctions. This occasionally results in unintentional humor. An illustration is the teen who was ruminating patriotically on why he likes being a (caucasian) American, and then finally exclaimed:

> It's like, yeah, we got some problems, and yeah, society is not what it was whenever, but it's like - you get free education, you know, most of the time you got plumbing. I mean, people - they take tv, the toilet, and like, vending machines for granted!

I am not sure that television is usually grouped with indoor plumbing and convenience foods in quite this way, although perhaps some would say that it truly puts the cultural role of television in perspective. I found that his explanation, while humorous, also celebrates the very themes of globalization and cultural imperialism that postmodern critics usually eschew. Such an unusual juxtaposition of ideas (from nostalgia for the pre-modern to the celebration of convenience, entertainment, and global capitalism), I contend, would be much less likely to emerge in an interview with an adult conscious of his or her self-presentation to a professional researcher. Thus research with adolescents not only provides the occasional unanticipated humor, but also can provide a freshness of

perspective that crystallizes both a common discourse and the teen's position in relation to it.

I begin this chapter with an explanation of common understandings of adolescence that shape this and other studies of this audience. I then review the media research and then religious research on adolescents to provide the foundation for the current study's research questions, contextualize some initial findings, and introduce the subsequent analyses of a case study.

Adolescence and U.S. Media Studies

Adolescence is commonly understood as a time of life in which both biological and emotional change occur as ties with parents are severed and the individual experiments with various possible social roles which may define one's adulthood. One of the earliest definitions of the term occurs in the writings of G.S. Hall, who in 1916 argued that the age-specific behaviors of individuals in this stage of life resulted from physiological drives. Adolescence as a *concept* which refers to this transitional stage in life, however, has a much longer history. Schlegel and Barry (1991) trace the existence of the concept to the preliterate western societal view of life as comprised of a series of stages. They point to the 6th century B.C. iconographic representations of the "Ages of Life" as evidence of the early concept of adolescence as a transitional stage between childhood and adulthood.

Adolescence is also believed to be the point in life at which persons begin to reflect on the meaning of life itself, and I take this to be one of the intriguing aspects of this stage in the "life course," as I will describe more

fully at a later point. It is this time when young people first begin to express a sense of themselves, although, as Gilligan and her colleagues (1990) have pointed out, this very "voice" is often suppressed among adolescent girls through our education system and social expectations.

Definitions of adolescence, like other issues requiring definition in this postmodern age of scholarship, must therefore be contextualized historically, economically, socially. Thus rather than seeking a totalizing definition of this transitional moment, I have adopted a pragmatic stance. Because my project required the social scientific exercise of limiting the sample by demographics, I adopt the definition of adolescence given by Elder (1975):

> Despite a lack of consensus among social scientists on the social boundaries of contemporary adolescence, the clearest marker for entry into adolescence is the transition from primary to secondary school (from sixth to seventh grade). Entry into one or more adult roles (marriage, parenthood, full-time employment, financial independence) is commonly regarded as the upper boundary (p. 3).

Yet I am also interested in moving beyond the concept of adolescence as a developmental stage and taking into consideration the adolescent's role within the larger society. This is a question more frequently raised in anthropological approaches to adolescence, and opens up questions of the role of adolescents in contributing to social change (See, e.g., Mead, 1928). My study foregrounds not the stage-specific behavior of young people or of developmental issues associated with this time period, but the patterns of discourse employed by teens, looking for points of connection and fissure between their own and their parents' approaches to the symbolic resources of the media.

Emphasizing the "upper boundary" of marriage, Modell (1989)

argues that changes in the U.S. concepts of adolescence and youth are inextricably linked to the sexual revolutions of this century beginning in the 1920s. Taking marriage as this central "rite of passage," he demonstrates how various factors, such as the extension of schooling into the late teens and early 20s, the declining sanctions against premarital sex, and the peacetime lack of military service for men redefined adolescence into a period that now extends into at least the mid-20s. Some, such as Giddens (1993), have argued that in fact today's baby boomers represent an "adolescent generation," continually seeking stimulation and self-satisfaction while experimenting with ever-new identifications and relationships. Yet for the most part, U.S. culture currently recognizes the time between approximately ages 11 and the early or mid-20s as a specific time in life with describable (if fungible) boundaries and contours.

Like younger children, teenagers have been the focus of some public debate and concern regarding the negative role of the media in their development. Frequently, these fears are supported with statistics concerning the amount of time teens spend watching television (20-30 hours per week, or an average of 9 hours a night on school nights),[1] listening to music (also 20 - 30 hours per week)[2] , or consuming other media (4 1/2 hours per week on the Net).[3] Over the past century, researchers have explored teenagers and the negative effects upon them caused by the movies (Blumer, 1933; Berkowitz & Rawlings, 1963), rock music and music videos (Greenfield et al., 1987), video games (Funk,

[1] Morgan, M., Alexander, A., Shanahan, J. and Harris, C.. (1990) report that children in grades 7 - 9 watch about 30 hours of television per week while children in grades 10-12 watch 20. For breakdowns of hours watched on school nights, see Snyder & Hoffman, 1993, p. 13.

[2] This statistic is given in Brown, Klein & Walsh-Childers, 1990.

[3] A CNN/USA Today/NSF/Gallup Poll (1997) reported this finding as a result of a national survey of persons aged 12 - 17.

1992), and television (Greenberg et al., 1993; Pearl, Bouthilet, & Lazar, 1982). These studies are based in social learning theory which posits that children will learn and then imitate behaviors they observe are rewarded (Bandura, 1977). A key implicit assumption of this type of research is clear in this statement by Strasburger (1995), who argues for the urgency of his research into television's deleterious effects on adolescents. He writes: "young teenagers may be the most susceptible population to television's hidden themes and messages because their identities and attitudes are evolving and more malleable (p. 12)." This concept of the "malleable" identity of teenagers has been influential in placing concern for media effects on the public policy agenda. The fear of violence on television, for example, becomes a debate that substitutes for the larger fears surrounding social, cultural, and demographic change - and how such changes might impact the dominant middle class. Television thus becomes a scapegoat for social/structural problems (Rowland, 1983; Jensen, 1990).

Some of the most widely recognized research into media and its "effects" on young people has been written by those trained in psychology (See, e.g., Zillmann, 1971; Berkowitz,1964; Singer & Singer, 1981; Comstock, Chaffee, Katzman, McCombs, & Roberts, 1978). In part, the popularity of this research rests with the public attraction to seemingly logical explanations to difficult social and material problems which then suggest answers to such problems. Encourage parents to control their children's viewing, or support legislation that restricts programming for young people, and the problems of violence in society might be solved, they imply. Of course, the concern for teens and the possible negative effects of media upon them is related to the tendency to see childhood as a

"protected space" - a culturally and historically specific approach to children which has varied in relation to the economy.[4] This view has had a long historical tradition, as illustrated in these words from Plato (1987):

> And the first step, as you know, is always what matters most, particularly when we are dealing with those who are young and tender. That is the time when they are easily moulded and when any impression we choose to make leaves a permanent mark...Then it seems that our first business is to supervise the production of stories, and to choose only what we think suitable, and reject the rest. We shall persuade mothers and nurses to tell their children, and by means of them to mould their minds and characters which are more important than their bodies. The greater part of the stories current today we shall have to reject (pp. 337-8).

I agree that some anxieties regarding the environment in which children are raised are valid, for while the media cannot be parsed out as *causes* of certain behaviors, they are a small part of the wider cultural environment in which young people live.[5] Yet it is also true that concerns about children and their welfare often serve as a symbolic expression of the fears surrounding modernity's impact on the society as a whole. Moreover, it is clear that psychological models of youth and development have been quite influential in the understandings of the relationship between media research and adolescence, as I noted in chapter 2.

Cognitive and Developmental Psychology and Teens

As a "life stage," adolescence has largely been viewed as the province of psychology, although as we will discuss, sociologists and historians have contributed equally important insights into this

[4] For a review of the cultural and historical significance of the idea of the "protected space" of childhood, see Coffield, 1987; Buckingham, 1993. For an example of the argument that children need to have protected space, see Meyrowitz, 1985.

[5] Piper (1994) makes an interesting if overstated argument for the role of media in reflecting the cultural values of thinness and submission which then may be said to play at least some role in reinforcing eating disorders and self-mutilation among teen girls.

transitional moment in life. Still, developmental psychologists tend to approach adolescence as a latter stage of childhood, such as in the writings of Piaget (1965, 1969, 1985) and Erickson (1965, 1968), who are perhaps the two most influential theorists in relation to research on adolescence. Piaget argued that children reached progressive stages of development through the acquisition of certain skills, which tended to be associated with chronological age. Between ages 2 and 7, children are in the *pre-operational stage* of cognitive development, as the child develops the ability to recognize symbolic representations of concrete objects. Between ages 7 and 12, children enter the stage of *concrete operations,* as they are able to understand numbers and the relationship between concrete objects and events. The final stage of *formal operations* is entered at approximately age 12, as the young person is able to employ abstract reasoning.[6] Wartella (1980) has used Piaget's schema to discuss how children interpret television images at various ages, noting that they develop greater means to utilize the media as they grow older. Rosengren and Windahl (1989), in their study of the uses to which children put media and the gratifications they gain from them, argue that children have different media needs based upon the developmental stage which they are experiencing at the time. Inserting media practices into Piagetian theory, they argue:

> During the course of cognitive development children thus become less dependent on their immediate perception, and their ability to deal with multiple dimensions of phenomena expands...We learn from cognitive stage theories how children obtain more and more means and resources for utilizing the media (p. 57).

While Piaget has been important in the research into children and

[6] The moral reasoning skill of young children, built upon a similar schema of cognitive development, is outlined by Kohlberg, 1958.

television, Erickson's theories of ego development and identity crisis have been an important foundation to how the relationship between teens and the media have been understood. Erikson, basing his argument on developmental stages, theorized that ego identity emerged in the fifth - the adolescent - stage of psychosocial development. To achieve this stage, an individual has to have a sense of emotional and physical well-being. In the teenage years, he claimed, people make decisions regarding self-definition that are irreversible and serve as commitments for life. These choices are made as the teen attempts to find some congruence between what is increasingly recognized as the larger context in which one lives, and how he or she conceives of his/her "place" within that context. As individuals recognize the need to find this place, they experience an "identity crisis" which then results in the discovery of their place within the world.

In his studies of adolescent uses of media, Roberts (Fine, Mortimer, & Roberts, 1990), building upon Erickson, has argued that adolescent identity crises are resolved as the individuals draw upon information from their environment, which includes the mass media. He argues that the media help adolescents to understand the choices which confront them regarding social roles, offering a way for teens to vicariously "role play" or experiment with social roles outside of any material consequences. Thus, media satisfy the needs for ego development and fulfill an important function in the life of teenagers.

One of the problems with the what has been called the 'uses and gratifications' approach to media analyses which have adopted these psychological perspectives as represented here is that they tend to tie

"needs" of individuals rather unproblematically to stages of development without considering other reasons people might be motivated to watch or listen to certain mediated programming. In general, researchers in this paradigm also do not question where those "needs" come from, except to say that they are universal and therefore, it might be assumed, not necessarily related to specific sociohistorical conditions. The question of whether or not society actually *allows* for the meeting of needs, particularly for those who are disadvantageously positioned socially and economically, is therefore not addressed in this framework, and thus all adolescents are treated as equivalent beings.

Some critics have argued, and I agree, that another problem with adopting Erickson's theory to the study of media and adolescents is that Erickson's descriptions are rather specific to the white, middle class and male experience (Gilligan, 1982). Marcia and Friedman (1970) note, for example, that females are socialized differently from males, and are more often rewarded for seeing themselves in connection with others rather than separate and independent from them. Adolescent women - and since Erickson's writing, increasingly their male counterparts as well - receive conflicting messages regarding their "place" with reference to both family and professional lives. Moreover, as Slugoski and Ginsberg (1989) point out, Erickson's theory posits a "benign" society which offers people seemingly unlimited opportunities within the social order. This implicitly locates responsibility for realizing opportunities (e.g., finding one's place in the social order through identity achievement) with the *individuals* rather than with the society. It is clear that not all young people have the same set of opportunities, nor are there desirable

professional positions for all who would want them. This, then, means that some individuals undergo identity crises only to discover that there is no "place" for them in society, due to racial, economic, educational, or other barriers.

There is another point within Erickson's theory that I believe is problematic, and that relates to the theory of stages of development. While I would not dispute the point that teenagers indeed undergo crises of self-definition, I am not convinced that adult identities are more "fixed" and "final" than teens as a result of that experience. Teens and their parents, I have found in this study, anticipate the adolescent years to be a time of testing and rebellion. While they might credit psychological stages of development for this, it is also possible to offer a cultural explanation. Bellah and his colleagues (1985), for example, argue that the expectation that young people must break away from family, home, and tradition in order to "come of age" is rather specific to western cultures. They relate this to the influence of early American child-rearing practices which emphasized the importance of raising self-reliant children for democracy.[7] Yet as Lasch (1991) points out, the project of "coming of age" or "finding oneself" is no longer limited to the adolescent, but is a part of American adulthood, as well. Our sense of self is continually changing, and the increasing interactions with symbolic resources available through travel, education, and the media of the latter 20th century suggest a need to rethink the concept of the self as fixed and impermeable. This is not a process limited to the middle class experience, but one that can be found across various social and economic categories.

As noted in the introductory chapter, I contend that adolescents are

[7] Bellah and colleagues cite Locke (1968) on this point.

an interesting focus of study for several reasons, apart from any unique experience of "identity crises." First, because of the cultural sensibilities which echo Erickson's theory, adults and teens alike are able to draw upon public discourses of identity-formation in sharing their own stories. Whether they are experiencing life changes as a result of psychological or sociological factors or a combination of both, young people and their families experience the adolescent and young adult years as times of flux. They are therefore particularly interested in and articulate about identity construction. Second, there are methodological benefits to studying adolescents. I am able to talk with adolescents while they still reside in the home of their primary parent(s) or guardian(s), therefore drawing upon the narratives of family members - particularly parents - in sketching out the various negotiations made by the teen in the attempt to "accomplish" an adult identity.

While psychological frameworks have been significant in research into teens and into the relationship of teens and media, the tradition of critical/cultural research originating in the work of the Birmingham Center for Contemporary Cultural Studies offers an important counterpoint. Here we find a model for a more contextualized approach to adolescent media audience research.

British Cultural Studies and Adolescents

In the early years of the Birmingham's Center for Contemporary Cultural Studies, adolescent identity and the media were a primary focus of concern, particularly with reference to the formation and maintenance of subcultural identity. A distinction between the U.S. and British

traditions is noticeable, as the influence of Raymond Williams, E.P. Thompson, and Richard Hoggart's culturalism, arising from a literary and historical rather than a social scientific or behaviorist tradition, encouraged scholars to explore media as a facet and reflection of social life rather than as something abstracted from social life which might or might not influence individuals. The particular constellation of symbolic appropriations adopted by a subculture served as an expression that reinforced that group's values and outlooks on life, as poignantly illustrated in Willis' (1977) study of working-class boys, *Learning to Labor: How Working Class Kids Get Working Class Jobs.* In addition to learning of how the media were used to reinforce the group's strategies, we learn how the acts undertaken by young men in their effort to rebel against their underprivileged social position - such as skipping or dropping out of school and resisting authority - ultimately reinforced their lower social status by structurally resulting in limited employment options.

Other 1970s-era work in the Birmingham school examined delinquent young people, drug cultures, fans of certain kinds of music, and stylish young people, all rooted in the Marxist questions of class conflict and resistance to the dominant culture (Frith, 1981; Hebdige, 1979; Hall & Jefferson, 1976).

Responding to the depth of research that working-class males had received in this tradition, McRobbie (1991) wrote essays on teenage girls in the British working class beginning in the 1970s, although famously collected and refined in *Feminism and youth culture: From 'Jackie' to 'Just Seventeen.'* In this work, McRobbie employs ethnography to explore the everyday lives of working class girls in south Birmingham neighborhood,

paying particular attention to the double standards the girls confront concerning norms of gender and class. Taking teen motherhood as an important symbolic moment, she explores the ways in which the girls' material conditions and gendered expectations limit their options, making pregnancy seem a viable means for reaching mature adulthood. She also concentrates on interpretations given to the popular teen girl magazines, *Jackie* and *Just Seventeen* , arguing that the meanings made were shaped by the experiences of either the middle or working classes.

Thus popular culture was viewed as a resource for self-expression and interpretation of social roles among adolescents. While this sounds similar to the project earlier articulated by Roberts, the differences between approaches is obvious in relation to attention to political/economic contexts. Hall and Whannel (1964) note about this process of identity-construction with media symbols, for example:

> The culture provided by the commercial entertainment market...plays a crucial role. It mirrors attitudes and sentiments which are already there, and at the same time provides an expressive field and a set of symbols through which these attitudes can be projected...Teenage culture is a contradictory mixture of the authentic and manufactured: it is an area of self-expression for the young and a lush grazing pasture for the commercial providers (p. 276).

Due to their interest in markets and Marxist theories, British scholars argued that particular symbolic combinations could comprise political resistance to the dominant culture, as Hebdige (1979) has described in his reviews of punk culture. Although this and other studies of the Birmingham school have often been critiqued, particularly by those in the U.S., for privileging class-based analyses to the exclusion of virtually all other categories, the introduction of semiotic, gender and race critiques by

McRobbie (1991), Hall (1985) and others have widened understandings of identity to include its relation to "difference."

The "Active Audience"

Some researchers have equated the British critical/cultural studies emphasis on symbolic appropriation with the "active audience" approach of the U.S. uses and gratifications traditions, as described more fully in chapter two. Yet as argued in that chapter, Ang (1996) and others have pointed to the differing epistemological traditions which result in different research projects and purposes. One example of the uses and gratifications tradition is the edited volume, *Media, Sex and the Adolescent*, in which Greenberg and his associates (1993) dispute the social learning or "effects" model to argue that adolescents actively participate in the interpretation of media content. Employing primarily (although not exclusively) the methodology of quantitative research, the various authors note the ways in which the different "variables" of age, race, gender, class, education, and developmental maturity influence teens' readings of media texts concerning sexuality. Their study shares some common ground with both the BCS findings and the current study in the implicit focus on individual identity and the relationship of the media to the cultural context. Further, like the current study, Greenburg and his colleagues are interested in the relationship of media tastes and meaning-making to cultural practices. Yet the approach to knowledge and the broad claims resulting from the research were much more consistent with the positivist paradigm of U.S. research than the more modest aims of critical/cultural studies.

Adolescents and Religion

As noted in the introduction, the current generation of teenagers is perhaps the least "churched," or experienced in organized religion, in several generations. Still, a high percentage of adolescents report a religious preference, and fully 95 percent claim belief in God or in a "universal spirit" (The George H. Gallup International Institute, 1992, p. 23). Nearly three-quarters of the teens in one survey of 13- to- 17-year-olds reported that they attended activities at their church or temple, a higher percentage than those teens reporting taking lessons or participating in extra-curricular activities (Moore, 1992). According to*The Religious Life of Young Americans* (The George H. Gallup International Institute, 1992), nearly 60 percent of teens claim a Protestant religious identity with over a third of these identifying themselves as Baptist (p. 53). Another 30 percent claim Roman Catholicism, while 2 percent claim they are Jewish, 1 percent claims "other Christian," and another 1 percent "other," while 7 percent offered no response to the question of religious preference. Nearly one quarter of the teens surveyed reported that they had watched religious programs on television in the previous 30 days, with the highest percentage of reporting from non-white teens and those from the South (p. 48).

Much the same as in the explorations of adolescence by media studies scholars, the Piagetian and Ericksonian stage development theories have had an influence in religious studies of youth as well. Using Erickson, for example, Parks (1986) outlines a schema of stages she believes adolescents must go through in order to achieve what she calls a "mature

adult faith." She writes: "Faith is, in essence, an act of imagination conditioned, in part, by structural development (p. 109)."

Perhaps the best-known of the theorists in developmental approaches to childhood and adolescence, however, is James Fowler (1981, Fowler and Keen, 1978).[8] Fowler (1981) argues that adolescence and young adulthood are the times in life when individuals develop from his "stage 3" or "synthetic/conventional faith" to a "stage 4" or "individuative/reflective faith." He contends that feelings of belonging are particularly important to persons in the third stage of faith. Persons in this stage are aware of having values and normative images in which they have deep emotional investments, but they have not made the value system itself an object of reflection. Thus, they tend to treat symbols literally:

> the symbols expressive of their deepest meanings and loyalties are not separable from that which they symbolize. Symbols of the sacred - their own and others - are related to in ways which honor them as unseparably connected to the sacred. Therefore, worthy symbols are themselves sacred. They are depths of meaning. Any strategy of demythologization, therefore, threatens the participation of symbol and symbolized and is taken, consequently, as an assault on the sacred itself (pp. 162-3).

Persons in this stage of faith development - the stage which, he argues, encompasses the majority of persons of any age involved in religious organizations - find the demythologization of symbols threatening and contradictory. He extends this group to include not only those who identify themselves as religious, but also those who identify with certain groups, such as labor organizations and others, that create a sense of collective identity much like a religious organization does. Thus, those

[8] For illustrations of the stage-related approach to religious development, see, e.g., Shelton (1983) and Wilcox (1979).

belonging to ethnic, racial, gender, or even sorority and fraternity organizations might be expected to dislike the inversion or demythologization of their symbols.

In the fourth stage of faith development, Fowler argues, individuals start to recognize that their beliefs, as well as those of others, are shaped by such factors as culture, social class, racial identification, gender, and economic conditions. Yet this realization must be coupled with "an interruption of reliance on external sources of authority (p. 179)." Such sources of "interruption" might include family members, religious organizations, and possibly fraternities and sororities. Persons must relocate authority in the *self* to reach stage four, which Fowler argues is a less frequent movement:

> By virtue of college experience, travel or of being moved from one community to another, many persons undergo the reactivization of their inherited world views and value systems. They come face to face with the relativity of their perspectives and those of others to their life experience. But they fail to interrupt their reliance on external sources of authority - and may even strengthen their reliance upon them - in order to cope with this relativity (p. 179).

Due to the reflexive stance toward belief systems coupled with personal autonomy for persons entering the fourth stage of faith, symbols are open to critical questioning. Thus symbols may be less powerful for these persons than they are for persons in stage three, but they are also opened for new interpretations: "Their meanings, now detachable from the symbolic media, can be communicated in concepts or propositions that may have little direct resonance with the symbolic form or action (p. 181)."

As is clear from the statement above, Fowler views college as a time in which young people are first introduced to ideas that might relativize and contextualize their own. Fowler's theory may be challenged on

several fronts, including its basis on hierarchical assumptions (which, like other stage theories, equate middle-class values and experiences with the "ideal" values of society). Yet the proposed differences between symbolic interpretation strategies among those with differing understandings of their faith commitment is intriguing, and we will return to this at a later point when we review the data analysis on the extent to which teens value religious authority, "sacred" symbols, and coherence in their own belief systems.

Another study that has several insights worth exploring in the current project is that of Mihaly Csikszentmihalyi and Reed Larson (1984), *Being Adolescent: Conflict and growth in the teenage years.* This study employed the Experience Sampling Method, which, while positivist in many of its assumptions regarding research design and data analysis, has the advantage of exploring *subjective experience* as opposed to only reported or observed behavior. In this method, research participants wear a beeper and are "beeped" randomly at several points during the day, at which time they are to record what they are doing and how they are feeling at that moment. The results reveal patterns of how and with whom teens spend their time. Csikszentimihalyi and Larson report, for instance, that teens generally find their time evenly divided between the home, the school, and "public" places such as places of work, retail outlets, or in the home of friends. They also not that teens spend the bulk of their time with peers: approximately a third of their time with friends and less than a quarter with classmates, in addition to nearly a third by themselves and less than a fifth with their family. This study echoed Bachman, Johnston and O'Malley's (1993) findings reported in*Monitoring the future.*

In this latter study, when asked which three activities they most enjoyed doing with other family members, 16 percent of teens named watching television. Although this was lower than the preference for outdoor activities, travel/vacations, and "just being together," it was a higher percentage than those naming talking, family meals, or going out to eat as favorite family activities. Attending religious activities was named by 10 percent of teens as a favored family activity, which was still higher than those naming going to movies, shopping, and partying/socializing/ entertaining (p. 194).

While not exploring the role of media directly, the Csikzentmihalyi and Larson (1984) study foregrounds "how much a person actually wanted to do whatever it was she or he was doing," or the amount of "intrinsic motivation" present in their activity (p. 49). Not surprisingly, they found teens to be *least* motivated in places most structured by adults and *most* motivated in places furthest from adult control. They note additionally that teens report being happiest when they were actively engaged in some activity, with sports and the arts or other skilled hobbies heading the list. It is in these contexts that teens achieve "negentropy," or what the authors eventually came to call "flow." As television viewing was often seen as a noninteractive and nonchallenging activity done with the family, it was less frequently named by teens as an experience of "flow" according to the authors' designations, and this has important implications for the current study which will be discussed at a later point. Yet the authors do report an example of how television viewing could serve as an experience of "flow," as in one case family members attentively viewed television with the intent of later reflection upon their common experience (p. 146). While

providing a number of helpful insights such as this, the study falters in its tendency to erase class, gender and racial/ethnic identifications as it builds a universal picture of teen experience. However, it shares one of the perspectives of the current study: rather than isolating media or other "external influences" on teens such as television, the authors contend that "growth in adolescence depends in large part on the amount of meaning one is able to extract from the culture (p. 281)." While my concern is less with individual growth and development than with changes in the social patterns of expression, I appreciate the fact that these authors locate *culture* as an environment and resource in the everyday lives of teens.

While a significant proportion of the research conducted on adolescence in the U.S. has been in the realm of developmental and cognitive psychology as noted earlier, there have been a few sociological and anthropological studies addressing adolescence which are worth noting. One is Robert Wuthnow's (1995) *Learning to care: Elementary kindness in an age of indifference.* Wuthnow reviews survey data on teens and volunteering and also interviewed several teens who were actively involved in volunteering. He argues that teens learn to care by being cared for and that volunteering experiences can help teens to translate their personal experiences of being cared for into a larger framework in which they come to see themselves as adults able to play a role in society. As he is interested in social change and the possibilities of igniting young peoples' impulses to volunteer through motivations that he argues are different from those of previous generations, he foregrounds not individual developmental issues, but those of social trends and how the individual experiences and expresses the social. In its questions, then,

his study has certain points of connection with the current examination of teens, their practices and their beliefs.

Religion, Youth, and Media

While the current project contributes new directions to the study of adolescents, media, and religious identity, it is not the first to address these three areas of interest. Consistent with the criticisms of Postman and others, those in the religious realm have addressed the "problems" of youth and media primarily in terms of the assumed behavioral effect of the media. In this section, I would like to review two of the more recent works in this area.

The first is *Dancing in the Dark* (1991), a text jointly authored by Quentin Schultze, William Romanowski, and other faculty members at Calvin College in Michigan. Although the authors insist that they have "no axes to grind or agendas to advance (p. x)," their conservative Reformed perspective guides their analyses and culminates in an argument for the employment of the Genesis account of creation as a foundation from which to criticize the role of popular and high art in culture.[9] Moreover, while purporting to speak of the universal experience of young people in North America, the text is biased toward the experiences of upper-middle-class suburban white youths. This is evident, for example, in the chapter on teen films, which celebrates the films of John Hughes as the "best" of the genre while ignoring altogether the films of Spike Lee, John Singleton, and other non-white filmmakers,

[9] The Genesis argument appears at the end of chapter 10, and states that Genesis offers a modelof "what God wanted for humanity" and thus subsequent critiques on popular culture "restores to art and aesthetics a central role for moral reflection and affirmation," pp. 274, 276.

as well as any films addressing themes other than those of the dominant middle class.

Rather than a sociological analysis of the contemporary situation, these authors offer a critique that is based primarily on a review of the content of popular culture and a history of a few relevant industries such as MTV, paired with unsubstantiated generalizations of both the media and young people. They argue for a completely different understanding of the audience than that which has been advanced here thus far, stating: "Most of the time audiences take in the offerings of the electronic media casually, giving almost no thought to their origin or impact (p. 48)." This leaves them vulnerable to the values the media will then "produce" in them (p. 48). As we discussed earlier, the assumption is that young people are especially vulnerable to the influences of the media. Young people, they note, are "not as interested in local or traditional ways of life as they are in the newest national and international fads and trends - "what's hot and what's not (p. 52)." To which I would ask, *which* teens are these? Research consistently demonstrates that young persons who are 'othered' by the media due to their racial/ethnic backgrounds or sexual preferences, for example, are very much oriented toward their local sources of support, which would include their families and religious organizations, and therefore presumably would find "fads" in the media possibly irrelevant or requiring negotiation within one's own context.[10] My research, while admittedly small in scope, finds in fact that this negotiative practice may be found among young people of *all* backgrounds. Yet often the authors make sweeping statements which implicate the media or foreground their

[10] One such study affirming this is Gillespie's (1993) study of immigrant families and their television practices. Whitlock (1989) notes the importance of local support for gay and lesbian youth.

influence; for example:

> To a great extent, the success of the media depends on the ongoing crises in social authority. It is within an unstable social climate of this sort, one to which the media themselves contribute, that more people turn to the media for authority and diversion (p. 58).

This then leads them to conclude that the media play an important, directive, and ultimately destructive role in identity-construction among youth:

> the electronic media often define youth's sense of time and space by their continuous influence on and substitution for the prevailing mix of local, regional, and national institutions...Put simply, youth feel more at home absorbed in the airwaves of national media than they do in their parents' house...In multiple ways, then, like a "double whammy," electronic media frustrate both the maintenance of local communities and the formation of adult identities (pp. 63, 65).

Unproblematically embracing Erickson's notion of identity as a stage-accomplished task, the authors charge that the electronic media delay the "formation of stable, mature identities (p. 65)." From this perspective, then, the authors set out a "call for standards," which would evaluate the moral messages of the media.

Despite its greater attention to media theory in the cultural studies tradition, Warren's (1992) book, *Communications and cultural analysis: A religious view,* arrives at very similar conclusions. Although more general in its focus, I chose to look at Warren's text because he has authored several books on religion and youth and devotes a chapter to the subject in this book (Warren, 1987, 1982). Again, his approach to audiences is quite passive, as illustrated here:

> Increasingly, the means of communication have been able to create for large numbers of persons a world of meaning they tend to consume, not to create or even to engage creatively (p. 125).

In an earlier section, he similarly argues:

> Should a people become dependent on the imaginations of reality produced for their consumption by unseen persons feeding them images, narratives, songs, and so forth, they lose a key feature of their humanization. They cease in some sense to be subjects naming and celebrating their distinctive life and instead become objects, with their life's meaning handed to them by others (p. 15).

While I agree that there are important distinctions to be drawn between popular culture as the culture of the people and as the culture commodified by mass media industries, I think Warren too quickly cedes the subjective position of audience members. I am more interested in how people continue to see themselves and their experiences represented in the mass mediated forms, and how even when functioning within the dominant culture's ideology, these symbolic forms provide resources for meaning-making. By assuming a "weak" subject in relation to "strong" media, Warren, much like the other authors, assumes the negative "power" of media in young peoples' identity-construction processes, such as when he desscribes a teenager who wants an expensive pair of jeans that her parents can't afford:

> Though the girl's anxiety was rooted in what her peers in school would think of her, she learned her basic lessons outside the school, through assumptions about reality communicated by advertising (p. 44).

There is no doubt that advertising plays a role in creating the "need" for the jeans; however, such a need can only become meaningful when it fits within other assumptions of the girl's peer and other social networks; the advertising has to resonate within the various discourses in which she is already positioned.

Both of these approaches to media, religion and young people would benefit, I believe, from a closer examination of the role of media in

young peoples' lives. Thus I hope the current analysis moves the discussion beyond *assumptions* of the media/audience relationship to interrogations of it.

Adolescents, Media, and Identity-Construction

In this final section, I would like to turn to the data generated during my first phase of research to reflect upon some of the various research examples reviewed in the previous sections of this chapter. First, as I have argued against an "essentialist" definition of identity, I do not believe that people are born with or develop a "core" of self that can be examined apart from their environment, and therefore I do not believe that the media can autonomously assist in or damage this "core." Yet I also am not convinced by the postmodern arguments that identity-construction is all "play" or "surface" and the media are but resources in this "game." I think we still need to account for the fact that people are rooted in specific historical, cultural, and geographical contexts that shape and inform the sense of self, as I have noted elsewhere. We need to account for the fact that, particularly in the current politically charged context, people resist having their identities reduced to categories or lightly dismissed as ephemeral.

Second, I also want to argue against the "powerful media" tradition, which portrays teens as duped and limited by media portrayals. The first and most obvious hurdle that this tradition confronts when talking to people of any age is what is termed the "third person effect;" that is, while teens readily acknowledge that the media have negative influences on *others*, they vehemently deny that it has any influence on they

themselves. Of course, people may not be aware of the role of ideological influences in their lives, so the *denial* of influence does not necessarily mean that there *is* no influence. However, it does point to the fact that teens are aware of themselves (if not of others) as critical viewers who bring at least some capacity to recognize and avoid overt manipulation (however they define that). The second challenge to this tradition arises in the recognition of the proliferation of mediated material that is now available. It is perhaps now more difficult than ever before to conflate dominant ideology and television content. While television does play a role in framing and reinforcing certain ways of viewing the world, it is fruitless to point to this single cultural industry as the *source* of an ideological system that clearly transcends it. This is why I believe it is necessary to see television within its broader cultural context of reception for teens, as I argued in chapter two.

Third, it is important to point out the dearth of programming containing overtly religious symbols and storylines in television and films, and yet the lack of complete agreement among audience members on what might be considered "religious" or "not religious" in programming. This mitigates against any attempt on the part of the analyst to define a clear boundary between "sacred" and "secular" media texts, as numerous individuals I interviewed mentioned "religious" themes they saw in such programs as *Star Trek, The X-Files,* and even *Walker, Texas Ranger,* while at least a few saw the more consensually agreed-upon "religious" programs such as *Touched by an Angel* and *Nothing Sacred* as "not religious" or even "sacreligious." Due to the dearth of overtly religious images such as church buildings and clerical

collars, however, it was not surprising to me that I found few teens who sought and incorporated religious information or "resources" from the media. Most of the teens were baffled, in fact, that I might see some connection, and several attempted to correct my "false" impression that there was *any* evidence of religion in the media. Even more young people were intent to demonstrate that the media had no influence on any of their beliefs. Due to the complexity of social lives, ideology, and other causal influences, it would be nearly impossible to argue definitively that teens' religious understandings were directly informed by the media or modeled after them. There were a few teens I interviewed who were exceptions to this in a certain sense, however. These teens sought out information on the supernatural that reinforced their views and employed media examples when constructing their religious identity in conversation with me - although of course I had suggested this construction, and thus had no evidence that it had occurred to them to use the media in this way prior to our conversation. These were teens who had rejected traditional religion and were intrigued by such mystical or paranormal experiences as ghosts, aliens, and demons. Amber and her father, who were described in chapter two, were one such case. Twenty-year-old Jodie, a low income caucasian adolescent sharing a household with the mother of her friend and four other young people, was another case.[11] When asked what television show was most like her own beliefs, she responded:

> It would have to be X-Files. Because, no matter what anybody says. My dad's a real science fiction freak, he's the one that kind of got me into that, thinking about aliens. Well, I've seen everything that everyone's compiled together about aliens. There's no doubt in my mind that we

[11] Other adolescents in Jodie's "family" had articulated this perspective, as well.

> are not the only intelligent life....God was a higher being. How do we know he wasn't an alien? On X-Files, Mulder, he would say something like that, 'how do we know God's not an alien?'

While her comments draw a connection between television programs and her beliefs (and also, incidentally, draw a connection between *her father's* beliefs and her own), I think they illustrate particularly well the tendency to *regenerate* religious beliefs within television texts. As described in chapter two, I use the term *regenerate* to refer to the way in which Jodie reads into the text a meaning that was not intended (at least, I believe that there is no evidence within the program that the character of Mulder has directly equated God and alien life). As noted in the previous chapter, I differentiate *regeneration* from Hall's *oppositional* reading, for Jodie draws upon what might be considered the *dominant* reading of Mulder as doubter of metanarratives/believer in alien forces for her projection of what he would say about God. As was the case for Joe and Amber, this is also not only a *negotiation* of meaning, as it moves beyond interpretation strategies related to the text. When she employs the example to explain to me her own views and thus construct within our conversation an element of her own religious identity, she is noting that the negotiated meaning contributes something to her larger belief system - even if what it contributes is less than coherent.

This actually then begins to call into question the placing of politics at the site of audience interpretation of texts. Numerous scholars have argued against the idea of the "resistive" reading of popular texts, and as I have already detailed some of these discussions, I will not go into more detail here. Yet the work of Csikzentmihalyi and Larson actually suggests a different approach to how we can conceptualize the issue of politics in

relation to adolescents: rather than focusing on whether or not the *viewer* is "active" or "passive," we should instead think about the fact that young people are rather passive about life in general and wish for the rare and fulfilling experiences of "flow." Csikzentmihalyi's and Larson's research points out that teens tend to see television as noninteractive and not challenging, and hence it engenders less "flow" than activities such as playing sports, skillfully working in crafts, or performing music. Yet even as these authors note, few teens seem aware of how to incorporate into their lives the practices that enable them to experience the "flow" feeling of being fully alive. Moreover, Wuthnow's research points out that while some young people seem able to reach a personally rewarding experience by acting on a concern for society, the link between the experience, the emotional reward, or the learned behavior of caring is not inevitable. This then introduces a different syllogism which seems to me an important precursor for any consideration of the entrance of politics into the issues of meaning-making and media interpretation strategies: do young people *want* to see society changed? Or do they instead want something to happen *to* them? I found much more evidence of the latter. As Lance, a 16-year-old caucasian complained, "Nothing very interesting ever happens to me." The matter-of-fact way in which many of the teens expressed similar sentiments illustrated for me the fact that teens do not seem to see themselves as actors (political or otherwise) who construct "happenings" in their lives. Several of the teens seemed to seek out*drama* in their lives, often through rebellious actions that they relished telling me about. But even those teens who were most sensitive to issues of injustice did not seem to believe that their actions could make a difference

in the wider society, and thus were indifferent toward expending the effort it might require. Once when I was talking about economic injustices with Michael, one of the "case study teens," I asked him if he ever wished he could have been part of the '60s civil rights movement that had been so influential when his mother was a teen. No, he replied, he didn't want to be a part of any movement; then, referring to his own aspirations, he said, "the only life I want to make a difference in is my *own.*" Along similar lines, when I asked about instances of racial or religious prejudice, several teens responded in individualistic terms that they had indeed experienced racism but "once they got to know me" the group with prejudiced views would change and thus the problem was solved - at least, it was for the individual.

Of course, as I have noted elsewhere, my sample of teens was quite small and therefore I am reluctant to claim that these experiences illustrate a larger trend, although I did see similarities between my own findings and West's (1993) argument that nihilism has become a central theme of the younger generation. Michael's nonchalant lack of a sense of responsibility for the wider society's injustices seems to me to be an outgrowth of what Bellah and his colleagues (1985) called *ontological individualism* : the belief that the individual has inherent dignity and that the individual is more "real" than society. Thus self-interest rather than civic or distributive justice is justified. Other teens similarly dismissed social action for practical reasons. Lance, for example, told me of how he disliked caddying at a racially exclusive country club because the golfers often made racist remarks. "I didn't like what they were saying, but I wasn't going to speak up or anything," he said, noting that if he hadn't

been caddying for them he might have said something. "But I wanted to have money. I wasn't holding that bag for nothing."

All this is to say that one of the key issues for understanding the politics in the media/adolescent identity question seems to me to be less about oppositional or resistive readings of media texts, than about the necessity of exploring the context of the therapeutic and individualist culture in which teens are coming of age and within which media texts are both produced and interpreted. In its narratives as well as in practices surrounding use, the media reinforce the idea that life happens *to* teens - or rather it happens to *some* teens: those teens that are attractive and well-dressed in particular. Clearly teens are "active" in this sense; as a colleague of mine put it, "if your subjects were so lethargic about life they all wouldn't be dressing the same and doused in CK One."(Shaw, 1997). Yet this is hardly the political activity academics endorse. Would it be possible for mediated messages to encourage political action and awareness among teens? Not unless those messages were also echoed in other contexts of the teens' lives.

While Bellah's notion of an increasingly individualistic culture is certainly an important facet shaping the adolescent experience, this is only one of several relevant changes wrought by late modernity. Part of the "task" of adolescence today, I argue, involves garnering the skills necessary to envision various possible outcomes to their actions in the context of what Giddens (1993) has termed a "risk society." This "risk" implies more than the increased exposure to new forms of danger, as he argues:

> To accept risk as risk, an orientation which is more or less forced on us by the abstract systems of modernity, is to acknowledge that no aspects of our activities follow a predestined course, and all are open to contingent happenings...Living in the 'risk society' means living with a

> calculative attitude to the open possibilities of action, positive and negative, with which, as individuals and globally, we are confronted in a continuous way in our contemporary social existence (p. 28).

Yet even as the U.S. has become a "risk society," the decline of the authority of adult institutions throughout culture in general has left young people with more autonomy, and hence more authority over their own behavior. Moreover, with the rise of part-time employment, young people themselves now have greater control over resources (both financial and educational) that allow them to choose the timing of the events in their own life course to a greater extent than in previous generations. This combination of factors results in a strikingly different approach to the future than the concept of one's "fate" which teens of earlier generations had been taught to accept, if implicitly. Perhaps in the past, teens felt that society held a specific place for them and their task was simply to find out what that was by undergoing an "identity crisis" of some kind, as Erickson (1968) postulated. Instead, with the rise of a plethora of potential courses of action, teens learn that they will, throughout their lives, continually be called upon to choose between "possible worlds." They have witnessed their parents and other adults in their lives changing their minds about mates, careers, and home locations, after all. Teens therefore have come to expect that while intimate relationships may offer fulfillment, such satisfaction may be ephemeral. They also expect that they will have autonomy and at least some measure of authority to make decisions concerning their own lives.

While there obviously are still struggles over authority and autonomy among teens and their parents concerning a number of issues such as respect for others and duties to the family, the area of religion

seems to be noticeably different. A number of parents expressed their interest in having their children "make up their own minds" about religion; the conversation between two single parent Mexican American Roman Catholic women in one of the parent focus groups illustrates this:

> Cathy: Now, Stephen has tested me with that, 'cause you know, he got to a point - 'Cause I would take him to church, and stuff, and then he's like, 'I'm not gonna go, I think I'm gonna be Moslem.' He told me that, and I go, 'fine. Go for it.' I go, 'whatever you want to do. You have got the teaching of the Catholic church, you think you can find something somewhere else, then, you're of age. Do it. But this is where I feel, where I'm comfortable with my religion here, and I believe in everything, all the teachings, and if you think you're not getting what you need, then you have to go where you have to.' And he stood about a couple of weeks, and then he was back.
> Nellie: See, now, I believe like this. I don't force my son to go -
> Cathy: You *can't* force them.
> Nellie: - Because I think, to me, church to me, it's not a punishment, it's supposed to be a blessing, and if you're forcing them, they're gonna be that much not wanting to go.

Parents are to lead by example, but ultimately they, too, reinforce the idea that young people will have authority to make their religious decisions for themselves. Raised in the turbulent '60s, the parents of today's teens approach religion, as Dykstra (1989) points out, with a certain "supiciousness, one that really can smoke out deceit, oppression, violence, and evil," that ultimately shaped their search for an "authentic" spirituality, religion, or relationship with God (p. 127). Yet this search, it seems, was inherited by the baby boomers' kids not as suspicion but as *cynicism;* not as a sense that there *might* be deceits but a *conviction* that there are. Just as the '60s shaped baby boomer angst, so the '70s and '80s shaped the current teen's sensibility that government, education, the economy, and even family life could fail them. They are even more skeptical than their parents when it comes to the culture's metanarratives,

to use Lyotard's (1984) postmodernist phrase. Their parents' search for self-fulfillment and authentic experiences of religion seems to strike some teens as a quaint and yet outdated notion.

The existence of this self-reflexivity, autonomy and authority then also has implications for how teens interpret symbolism, and religious symbolism in particular. First, while most of the teens I interviewed identified themselves as at least marginally religious, very few of even those most involved in formal religious organizations expressed a sense that their religious belief systems were coherent, or even that this might be a desirable aspect of their belief system. Second, while a few of the teens involved in organized religion seek out information on their tradition, most, regardless of background, expressed much more interest in "quasi-religious" supernatural happenings, such as ghosts, aliens, or paranormal encounters. Todd, a middle class caucasian 19-year-old who was raised Lutheran and now attends a nondenominational church, explained his beliefs in this way:

> I believe that, I don't know if there's a - . I don't believe in - . I believe in God. But I believe God is a higher power... I believe you choose your own destiny, but I also believe in fate...I don't know, I believe that there are - we're not the only intelligent life in the universe....And I think God is like a higher power, kinda like an energy, like a spiritual energy.

The interest in alien life and supernatural forces tended to be reflected in their choices regarding television programming, although not surprisingly, few of the teens were interested in programming that reflected the views they held that were associated with traditional

religion.[12]

My findings call into question some of Fowler's claims concerning adolescence and religion noted earlier. Fowler describes adolescents as usually in the "stage 3" of faith, noting that they are interested in belonging but, as I had found, they generally have not reflected upon their faith and its coherence. Yet Fowler also argues that teens in this stage feel that those symbols that depict their beliefs should be considered *sacred;* that the symbols are inseparable from that which they represent. Fowler argues that when the symbols held in highest esteem are denigrated in some way, teens in this stage of faith development would understand this as an assault on the sacred itself. I found, however, that only one teen in one of the focus groups expressed this concern, which was articulated as a criticism of *The Simpsons* as "sacreligious." When *The Simpsons* came up in other interview settings, many simply drew distinctions by genre; "*Touched by an Angel* 's more serious than *The Simpsons.*" "*The Simpsons* is funny," while others seemed incredulous: "There's religion in *The Simpsons* ?! What?!," as one teen asked. A number of teens deemed *The Simpsons* unacceptable based upon moral values or stereotypical portrayals, but no one except the one teen mentioned sacrelige or offense at any of the portrayals of religion in this or other programs. If they chose not to watch a program that had something to do with religion, it was because they felt the program was "geared for adults" (rather than teens), "boring," or not relevant to the "real-life" problems of teens. While all of the teens were familiar with at least a few examples of

[12] The only exception is Michael, whose family watches religious programming on the Trinity Broadcasting Network (an evangelical Christian cable channel) together. Michael was also the only regular viewer of *Touched by an Angel,* although several of the teens watched it less frequently.

the portrayals of religion on television, there was far more consensus that what was portrayed was close to their own beliefs than somehow sacreligious, which is what one may have expected given Fowler's typology. I believe this is a reflection, in part, of the lowered status of religion within culture in general. While religious institutions could once argue that they had the authority to decide how and when religious symbols might be used, there is now an increased tendency for religious symbols to be utilized in a way that is wholly separate from the institutions. Teens are used to seeing religious symbols in places other than in traditional religious settings - in music videos, on tattoos, in advertisements, in art galleries - and these occurrences are not usually meant to refer to the experience of the faithful community of which they may or may not be a part (except, perhaps, in irony). This disconnection of the sign-referent system has been linked to late capitalism's commodification of symbols and the regnant authority of the media institutions to use any of a culture's formerly "sacred" symbols in new combinations.[13] Ironically, this is especially true with the symbols of traditional Christian religions - the very ones with which Fowler was most concerned.

Moreover, as noted earlier, the experiences of young people today are very much shaped by increased autonomy as well as the relativized notions of truth articulated in everything from postmodern theory to secondary education's commitments to tolerance and pluralism. This changed environment for young people also challenges Fowler's stagewise analysis. Teens, like persons of all ages, see themselves as "bricoleurs"

[13] Baudrillard (1981) has made this argument. The commodification of religious symbols and the consequences of the declining authority of religious institutions over religious symbols is analyzed in Clark and Hoover (1997).

who can use or avoid religious symbols according to their needs. The symbols are not inherently sacred, but can be *made* sacred (or at least meaningful) based upon the context in which they are employed.[14] Thus, teens believe that their own religious belief systems and their own use of religious symbols, regardless of idiosyncrasies or inconsistencies, are as worthwhile as anyone else's. How teens come to this way of approaching religious symbols, therefore, has as much to do with the cultural environment as with their cognitive development. What makes a symbol meaningful is not the acceptance and reverence of the sign-referent relationship which eventually becomes relativized, but the symbol's usefulness in specific contexts of meaning. This, then, informs the analysis of the connections between religious symbols found in the media, meaning-making systems of teens, and their religious identity-construction processes.

This chapter has been an attempt to lay the foundation for the current study's focus on issues of adolescence, religious identity, and the role of the media in this process. Yet many of the insights noted here are best articulated in the narrative of Elizabeth, a fifteen-year-old girl whose project of religious identity-construction serves as an important case study illustrating how these various trends and observations work themselves out in the real life of one teen. It is to her story that I now turn.

[14] For an analysis of how this process occurs in the megachurch setting, see Hoover (forthcoming).

Chapter Five

Case Study: Elizabeth as a "Universal Singular"

In this chapter, I present a narrative biography of Elizabeth, a teen I have come to know quite well in the past year. I believe that such an approach affords the opportunity for social analysis beyond the individual case. Using Sartre's notion of the "universal singular," I see Elizabeth as both a unique individual and the embodiment of the social world that has reflexively produced her (Collins, 1980). By looking at her story, we learn both about how an individual constructs a life within "conditions not of her own choosing," and about the broader conditions which limit, shape, and give meaning to the ways in which she "speaks" certain widespread discourses of U.S. culture.

One of the obvious benefits of exploring the life of an individual is that it becomes clear immediately that no person is simply a sum of his or her class, gender, race, and place positions. As Mary Douglas (1970) argues:

> The physical experience of the body, always modified by the social categories through which it is known, sustains a particular view of society. There is a continual exchange of meanings betwen the two kinds of bodily experience (social and physical) so that each reinforces the categories of the other (p. 65).

Thus the individual serves as the site which mediates and grounds our understandings and experiences of the external world. As individuals *articulate* their individual identities in relation to wider discourses of the culture, the messages and symbols of the communication media may be expected to offer shape and substance to the individual's narratives. Geertz (1973) has noted, for example:

> In order to make up our minds we must know how we feel about things; and to know how we feel about things we need the public images of sentiment that only ritual, myth, and art can provide (p. 83).

How these various images are drawn upon and incorporated into preexistent discursive strategies is a subject for the analysis of the story. Thus after Elizabeth's narrative biography, I place the story in the context of the various discourses she evokes in her narratives of identity-construction. I hope to illuminate the ways in which Elizabeth is a product of various factors such as her class, gender, race, place, and history. Yet while she exists within these conditions which shape her view of the world and hence her discursive strategies about it, she employs agency within that world creatively and in some ways surprisingly.

Elizabeth's Story

Responsible and reluctantly independent, Elizabeth is the oldest child and only daughter in the white, lower middle class Farley family. I was interviewing and getting to know her during the years in which she was 14 to 16.

Elizabeth's mother Jean supports Elizabeth and her 11-year-old brother John with a part-time clerical job in a property management company, supplemented with child support funds, while attending community college on a part time basis. The Farleys had lived in North Dakota and in Montana previously, following Elizabeth's father's jobs in the upholstery business and as a prison guard. They had settled in the Denver area eight years ago, although her father's job as a radio tower inspector required a great deal of time away from the family - sometimes three weeks at a time - which placed an increasing strain on

the marriage. Elizabeth's parents were separated in the spring and a divorce followed rather quickly in the fall of 1995, just six months before I initiated contact with them.

In the beginning of 1996, the year following the divorce, Elizabeth and her brother John saw their father every other weekend. A job loss forced him to move from nearby Arvada to Dallas, however, placing another not-infrequent financial strain on the family. After yet another job loss, in 1997 he moved to Minneapolis to work in an upholstery factory once again. Elizabeth and her brother have seen him infrequently since he moved from the Denver area. They spent the summer of 1997 with him, living in the two-bedroom trailer he shares with his girlfriend. For the summer of 1998, however, Elizabeth's father stated that he can only afford to have one of them live with him. Thus while Elizabeth wants to give her brother John the option of deciding whether or not he wants to go, she hopes that he will choose to stay at home so that she can live with her father again.

Elizabeth lives with her mother and brother in a quiet middle-class neighborhood which has become increasingly upscale in recent years, evidenced by a large community recreation center and high-priced new homes recently built several streets away. Although her school district's affluence is mixed, most of her close friends live in two-parent households and have fewer financial concerns than Elizabeth's family does. Not surprisingly given the erasure of class in the U.S., however, Elizabeth is rather unaware of economic differences in her school. Once, for example, she told me that cheerleaders at her school paid nearly $200 each for their uniforms and warmups. When I asked if the cheerleaders

were mostly rich kids, she seemed surprised at the thought, and vaguely said that she thought probably the school would offer support if someone needed it. As she and her friends have reached their 16th birthdays this year, she has begun to recognize some discrepancies between her experiences and those of her friends, noting that while a number of them have received new cars for their birthdays, her mother is struggling to find the money simply to pay for her drivers' education and license.

Elizabeth has also reevaluated her goals of college education in light of her family's financial strain. While she had once planned on attending a four-year college and pursuing a degree in psychology, she now is planning to begin her college career at a local and less expensive two-year college. However, her grades in school have not been high enough to qualify her for merit scholarships and her lack of motivation in her high school classes does not bode well for her potential success in college.

Elizabeth's experience of class differences are not limited to her immediate surroundings; she is very involved in the Lutheran (ELCA) church her father's family still attends in an older and less affluent section of the city which she describes as "kind of low class" with "small houses" in what she says is "called the bad part of town." In addition, her mother Jean is involved in a weekly single parent support group comprised of mostly working- and lower middle class families that offers social activities for the whole family once a month or more, introducing Elizabeth to more peers who share her family's financial struggles. While Jean's identification with this group is quite high, Elizabeth notes that her church has been the primary source of support for her through her parents' divorce. Conversely, Jean attends the church - which her former

mother-, father-, and brother-in-law's family attends - only infrequently.

Elizabeth volunteers as a leader with the church's weekly after-school program for children and occasionally acts as a communion server in the Sunday morning worship service. She explains that she attends the church primarily for its youth activities, however, as she likes the youth director and the older kids involved in the group. When the youth director announced that she "wanted to spend more time with the grandkids" and so was giving up her volunteer position, Elizabeth was saddened because no one maintained the youth program. However, she is sympathetic to the church's emphasis on others. "The church's priorities are really with the older people," she explained to me, which seemed evident by the large and active senior center I observed upon visiting the church. But when I asked whether she would consider joining another church which had more young people, she smiled peacefully and said she wouldn't join a different church because she's "too connected" to the one she attends now. She has occasionally attended other churches with friends, but enjoys being in a smaller group with people who are not from her high school.

The Lutheran church Elizabeth attends is largely traditional in its worship style. Elizabeth's mother Jean would prefer a nondenominational "easygoing, new age" church like one she attended when she was in her early twenties, as she describes it. However, she is content to support Elizabeth in her involvement, and she and Elizabeth's brother joined the Lutheran chuch during the period of my study. Jean's own growing-up years were spent in a Southern Baptist church which she rebelled against in her teens. Jean's parents and sister, all of whom Jean

sees regularly but Elizabeth and John do not, have since ceased their involvement in the Southern Baptist faith, as well. None of the Farleys are involved in political or civic activities of any kind, nor are they involved in other regular activities beyond the church and support group. Elizabeth, however, has begun to volunteer as a coach for a young children's neighborhood soccer club.

Family Media Practices

Due to her mother's busy work, school, and social schedule, Elizabeth often has responsibility for watching her younger brother John, a task she accepts unquestioningly. Thus the siblings are usually home together in the afternoons and a few evenings a week. They often watch television together in the afternoons, although while John watches cartoons and action movies on television throughout the afternoon and evening, Elizabeth prefers to spend time in the evening doing her homework, talking on the telephone with her friends, reading teen magazines or horror novels, or listening to music.

When Elizabeth's parents were still married, the family gathered each Friday night to watch the TGIF lineup on television (a series of family-friendly sitcoms), a ritual both Elizabeth and her brother recall fondly. The divorce naturally brought an end to this weekly ritual. On the weekends when they were with their father, Elizabeth explains, they were usually still in transit or eating dinner when the shows were on, and now their mother prefers to take them along with her on social activities on Friday nights. In addition to this change in practices following the divorce, Jean reported that she watches television much less now than she

did when she was married, preferring instead to read self-help and fiction books or attending social events. Elizabeth's media practices did not change much after the divorce.

A Teen's Favorite Medium: The Phone

Elizabeth's best friend attends her former school, so their main contact is a daily phone call. As she describes their relationship:

> Elizabeth: I call my best friend about every day. We got really close at the beginning of 7th grade, we started calling each other every day...
> Lynn: What made you get close to your best friend in 7th grade?
> Elizabeth: I think it seemed more like we were going through the same thing, the shock of going into junior high and changing every single class, all the homework and stuff. It really changed everything, your perspective on life, and stuff.
> Lynn: So how'd you start talking with her every day?
> Elizabeth: Um, for a while there was this really cute guy at school and I used to talk about him and she'd talk about the guys at her school, and so on. That's probably where it started. 'Cause there were so many more new people, and [we talked about] what we thought of those people.

Elizabeth notes that while she and her friends frequently talk about "cute boys," they also talk about

> other people, or if we like this class, or what we're doing in that class, other people at the school we like or don't like, the teachers. We like to really gripe about the teachers. We gripe about our parents and the family stuff we have to go to. What a pain in the butt it is to have to do this or that.

They also use the phone as a means of getting support. Elizabeth notes that when she first learned that her parents were getting divorced, she called each of her friends to talk about it with them. Because of the limited mobility of teens who are unable to drive to places on their own, the phone becomes an important way to maintain relationships outside the home.

While on the phone, Elizabeth and her friends also listen to music and play on the computer, according to a discussion they held together on the topic. They consider it rude when someone is watching television while talking on the phone, as television can serve as a distraction. Homework can, too, although the primary problem with attempting to do homework while talking, they acknowledged, was that the homework does not get done. The exceptions are the calls in which homework is discussed among the friends who are in class together. Likewise, television is an acceptable side activity when the two conversants watch the same program at the same time, a practice Elizabeth and her best friend occasionally engage in. "We'll watch *90210* together, we'll make fun of the actors," Elizabeth says. They also hate it when their calls are interrupted by either a parent or sibling who wants to tell them something. They fear that the family member might overhear something, or worse, might embarrass them in front of whoever is on the other line.

There are no rules limiting Elizabeth's use of the phone (they have voice mail, caller ID and call waiting). Although Jean noted that she occasionally has had to restrict the amount of phone use as a punishment, Elizabeth related only one story of restriction, which actually was related to the caller rather than an imposed punishment. One of Elizabeth's former friends had joined a gang and began coming to Elizabeth's home in the afternoons when her mother wasn't home, primarily, Elizabeth believes, to use the telephone to be in contact with her fellow gang members. While Elizabeth was uncomfortable with this practice, she was unsure as to how to end it. Then, one of the gang members called Elizabeth:

> Elizabeth: Once, this guy called and he was 16 (Elizabeth was 14 at the time) and he's kind of involved in a gang, and he goes to Shaw

Heights, and my mom got kinda freaked out about that because she knew that he was in a gang. So she didn't give me his phone number and I didn't have it, so.
Lynn: How did she know he was in a gang?
Elizabeth: His cousin or friend was going out with my friend for a little while, and he got my number from a friend. He just called, and I was like, 'Whoa.' He just called to say 'hi' and my mom had a fit. She thought I was gonna get involved in gangs or drugs and stuff.

Since Elizabeth was forced by this event to reveal her friend's practices, Jean forbid her friend to come over to use their phone in the afternoons. Relating the story to me several months later, Elizabeth seemed relieved that her mother had ended this difficult situation. While she regretted that her friend had joined a gang, she did not regret the end of that relationship, as she felt that her friend was using her to gain access to a telephone.

Since that time, another of her friends has begun to get involved with gangs. She has used the phone to distance herself from this friend, "screening" and then not returning her calls. She's hoping this friend will "take the hint" that she's no longer interested in having contact with her. She feels badly about distancing herself because she knows she is one of this girl's closest friends, but she says she does not want to have anything to do with gangs.

Print Media

As noted earlier, Elizabeth spends a lot of time reading teen magazines. She subscribes to *Sassy* and *YM*, and also occasionally purchases *Teen Beat* and *Tiger Beat* in the stores. While *Sassy* has "more fashion stuff," she prefers *YM* because "it is centered around what you do in school. And *YM* has a lot of stories and articles that the readers have

written, or that are more appealing to what the readers want to know about." She especially enjoys the advice columns, and the "Say Anything" column in which readers write in about their most embarrassing moments. She has even used the magazine in her efforts to give advice to friends:

> Elizabeth: A couple months ago they *[YM]* totally focused on sex, and having safe sex, holding out or - . They did an article on teen pregnancy and teens that have children. I knew this girl that was pregnant for a little while, she had an abortion. And I let her borrow that article. I think that kinda gave her some emphasis on what her decision should be and stuff.
> Lynn: You mean you think the article encouraged her to think about having an abortion?
> Elizabeth: Yeah.

She also cuts photos of male celebrities out of magazines and hangs them on her wall or offers them to her friends. While she enjoys reading about the female celebrities, she notes that she never pins up their photos:

> I always keep the pictures, but I never hang 'em up. I hang 'em up for a little while, and it's like, there's one female picture and these billion boys. It always seems out of place. So I always end up taking it down.

The only other occasion that has prompted her to remove a celebrity photo from the wall involved a time when she learned about the celebrity's sexual preference:

> Elizabeth: Um, I used to be kinda into Keanu Reeves. But, I don't know, after a while I started hearing rumors about him and it's like, okay, that's not cool. So I took down all his pictures.
> Lynn: Like, what wasn't cool?
> Elizabeth: I don't know why, but they said he was gay. I don't know why. They just did, and it was like, okay. He doesn't seem to be that interested in females, so I guess I'll just take this down now. (She chuckles and pauses). But I think if it was that same thing with Devon Sawa, I don't think I'd care. He's too cute.

Dating on the 'Net

Just three months after the divorce was finalized, the Farleys introduced an important new media practice into their lives with the purchase of a computer and modem. Elizabeth and Jean have spent quite a bit of time experimenting with chat rooms available through online carriers. In fact, they creatively rotated through the free hours offered by Prodigy and Microsoft Network over a span of about six months. While they have the introductory AOL disks as well, Jean is concerned that even the free time comes with too many fees, making it "too expensive." They currently are members of a free net service which Elizabeth describes as, "it's not a good address," although they are vigilant in looking for opportunities to have another service provider through a free offer.

Elizabeth's primary use of the Net has been to communicate with other young people in youth chat rooms. While she has explored Christian youth chat rooms, she found that their smaller size meant that she frequently was unable to participate. Moreover, her experience was negative:

> When I did try to get it, it was full. Or, I just looked in there, and nobody was talking, they were just sitting there. Every once in a while, someone would pop up and say, 'Okay, who's going to talk now?' Nobody was saying anything, so I just decided to leave. It didn't sound interesting to me.

Her chat room and follow-up e-mail experiences have afforded her an opportunity to experiment with male/female relationships which she had lacked in her junior high and high school. "I'm not too popular with the guys," she'd once explained to me. However, on the Internet, she was "dating four guys at once" and seemed to have no difficulty meeting and developing relationships with boys. When I asked her if she felt that it

was easier to meet guys on the Net than in real life, she replied:

> I think it is. Because you can be whatever you want to be, and the guys can be whatever they want to be. So it might not necessarily be an *honest* relationship, but it's fun. Because you don't get really serious, because, obviously you could easily get involved with a guy where you could actually talk to them and see them than on the Internet. So I think it's just for fun.

One of the "fun" aspects is the opportunity to learn from others, while another is being found desirable as a conversational partner:

> You can get on there and ask for, like, get a survey. 'How many of you people like doing this?'...And you'll get twenty answers like that...You're like, 'are there any guys that want to talk?' You go into a room of a hundred people, you go, 'are there any guys that want to talk?' And instantly you get like fifteen people saying, 'yeah.'

As there are fewer females than males in these youth chat rooms, Elizabeth speaks of holding pseudo-contests in which she has offered, "Whoever can get to this [chat] room first wins" the opportunity to talk with her exclusively.

Not surprisingly, the boy/girl relationships which occurred on the 'Net seemed to have a different tone than the often hyper-aware and jealously guarded relationships of "real life" high school. She explains, for example, that she once introduced one of the guys she was "dating" to her friend Susie:

> After like a minute, he asked her out. At first I was mad because he was mine! But then I decided, well, I have three others so I should just let her have him...One of the other three was his brother, anyway.

Dating two brothers in the "real world" would be considered pretty unusual, she concedes. While she and her mother rotated through the free offers of online service, she would e-mail these former "dates" to continue the conversation, although frequently months would pass

between contacts. Once, she said, she had to "refresh this guy's memory" by sending him old e-mail notes she'd saved, which then got them back into conversation.

What exactly is an Internet "date"? Based on Elizabeth's descriptions, it seems to me that any kind of meaningful and exclusive contact between a male and female, no matter how ephemeral, might qualify. She explains the process of "dating" in this way:

> What would usually happen is that we would meet for the first time in a chat room, and then if I decide I want to talk to them more personally, I would get a chat room for only like 2 or 3 people, so we wouldn't get people coming in and out all the time. And we'd talk for a little while, until one of us had to leave, or something. We'd exchange e-mail addresses, and we'd like write every once in a while. And like, we could get together at a certain time. I'd say, 'I'm gonna be on the 'Net at this time, if you can meet me at this chat room at this time, then I'll see you there.' And if they can't, then that's okay. So it's kind of more lenient. And, they never know if you're dating other people or not. So you could be extremely serious, and not know it, or you could be dating other people and not know it.

Internet "dates" and the boundary between dating and friendship seems more fungible and ambiguous, therefore. In fact, while two of Elizabeth's long-term "dating" relationships had been "romantic" at the beginning, they have evolved into friendships over time.[1] They seek each other's advice about boy/girl relationships in which they're engaged at their respective schools. Elizabeth says that with one of them, she has also become more of "herself" over time:

> I'm kind of more friends with him, 'cause he's also really shy like me, so I'm more myself around him. 'Cause I think when I was on the Internet with him [at first], I was more assertive and stuff, like I would want to be more spur-of-the-moment, but then when I started e-

[1] "Getting romantic" consists in checking if your conversational partner has a boy- or girlfriend and then continuing the previous conversation.

mailing him longer, it became more what I was really like.

The other "date," however, still does not know the "real" Elizabeth, in part because she enjoys the "fun" aspect of re-inventing herself for him, but also because she is less trusting that the relationship would continue with more honesty:

> The guy who lives in Oregon, I'm mostly like myself, but occasionally I'll be, I'll say, like with the last [school] play, I'll say I was in it, when I really wasn't. I was like in the crew or something. But I was around it enough to know the lines, and stuff. So I was like, 'Have you ever been in a play before? This is my first play, and I'm really nervous.' And he buys it!

She also has told this "date" that she is older than she really is.

She expresses no interest in meeting these "dates" in person or even in conversing with them on the telephone, however. In fact, when one suggested that they talk on the phone, she deliberately kept her phone line busy during the appointed time so that he would not be able to get through. She says that they didn't "talk" on line again after that, something she seemed to have no regrets about. When another "date" became "too serious," she stopped talking with him. She also notes that while she had never "met" anyone online from her own school, she did meet someone from a neighboring school once and promptly terminated the relationship:

> We started comparing notes about who we knew in each others' schools. But I didn't want to meet him, or someone from my own school, because then what if I knew who he was in person and he said something mean about me, I'd be like, hurt.

"Dates" with faceless and voiceless boys from faraway places held no such possible consequences. Clearly, it is easier for Elizabeth to participate in such relationships. The anonymity of the 'Net seems to have given

Elizabeth some power that she didn't have in the "real-life" situation of dating. This power comes from the absence of the strictly guarded boundaries and gendered expectations of looks and popularity in high school. The way she describes her constructed 'Net persona reflects these boundaries and her ability to subvert them through her verbal presentation:

> Elizabeth: You don't really see them in real life, and you can't really meet them, and you can be however you want to be, and you can be really assertive when you're shy, like me, I get more assertive on the Internet. And you can just change whatever you don't like about yourself. And you can just meet people, you can meet them once and never have to talk to them again.
> Lynn: How do you describe yourself on the Net?
> Elizabeth: Usually I describe myself skinnier, or taller. Skinnier and taller, with longer hair, and a lighter blond usually. And usually I act a lot more aggressive when I'm on the Internet. I just express my feelings a lot more in the chat rooms and stuff, so if somebody talks about something I don't like then I'll say it. And I would probably never do that in class, in school and everything.

In "real life," Elizabeth describes herself as "more passive" than on the 'Net. As seen in these comments, the gender-defined roles of passivity, receptiveness, and perhaps the norm of females making a greater emotional investment in relationships than boys could be subverted on the 'Net with fewer possible consequences of rejection. Moreover, because of the reduced emotional investment, relational breakups are less painful. As Elizabeth notes, "'If you break up with that guy, then you can just go out and find another one."

There can be moments of confusion that result from this ability to self-consciously construct the self, however. Elizabeth related an experience when both she and a girl with whom she was talking assumed that the other was a male:

> I thought she was a guy, and she thought I was a guy. So we went in, and we started talking, and she goes, "Oh, I'm a model for a teen magazine." And I was like, "No way!" And she said, "Yeah, I'm gonna be in next month's." And I looked at it, and there was a girl on the front cover, and I was like, wait a minute! So it got me weirded out, and I got back on the Internet, and I said, "Are you a girl?" And she said, "Yeah, is there a problem with that?" And I said, "Well, I'm a girl, too." She goes, "Oh my gosh! I thought you were a guy!" So sometimes you can get kind of confused if you don't specify who you are.

While Elizabeth has communicated with girls on line before, in this case of mistaken gender identity the relationship quickly ended.

Elizabeth notes that on line, boys are much more open and vulnerable than in real life, which seems to be part of a role reversal regarding who is in control emotionally in Elizabeth's relationships:

> You know, when you're online, and you're talking to a guy, just before you leave, nine times out of ten, they'll say, 'I miss you, kiss, kiss.' Or something really corny like that. But they won't say that on the phone.

In a 'Net relationship, Elizabeth is able to observe the comments of her male friends with greater distance and skepticism than is possible in "real life." Perhaps this is due to the reliance on words rather than the more easily manipulated combination of words and body language, and it may also demonstrate Elizabeth's greater facility with words when compared with her male counterparts. Instead of relying on the standard high school currency of looks and popularity, then, Elizabeth can parlay her good verbal skills and sense of humor into control of the relationship on the 'Net.

Yet while Elizabeth's constructions on the 'Net are creative and allow her to seemingly transcend her physical self, these constructions do not transcend existing social expectations, nor are they resistive of them. After all, her re-construction of self is quite conventionalized within the

cultural expectations of the "ideal" female appearance. It is this ability to construct herself as ideal that empowers her to be more assertive and therefore challenge the expectation of passivity and accomodation which she more readily adopts in "real life." Thus her employment of the technology of the Internet is in keeping with social conventions concerning gender roles. Similar arguments against the potential new communication technologies hold for subverting social systems have been made elsewhere (See, e.g., Rakow & Navarro, 1993; Rakow, 1988.

Discussion: Media Practices

As discussed in an earlier chapter, the project of identity-construction is very important in the life of U.S. teens, and they are quite conscious of how they use various media in this effort. Elizabeth draws upon the media both consciously, as in the case of the teen magazine which became a vehicle for expressing her support for the controversial issue of abortion, and less consciously, as in her use of the phone to garner support from friends, to passively communicate a desire to terminate a relationship, and to avoid a more serious confrontation between herself and gang members. She also uses the media to affirm appropriate gender behavior, expressed in her discomfort with hanging pictures of females on her wall of celebrity pin-ups, and her discomfort with Keanu Reeves' picture after discovering he was gay. It is worth noting, although not surprising, that all the male celebrities she pins up are white. One of the temporary female pin-ups, however, was the African American rhythm and blues group TLC. The celebrity wall serves as an important identity marker for Elizabeth, giving her an opportunity to express both her

aesthetic preferences while excluding those whose views and experiences which do not fit with hers.

Not surprisingly, Elizabeth's media practices differ from her mother and brother. It is also not surprising that she draws upon some of the same discourses as her mother when discussing the media and its role in her life. However, what is interesting is that there are several points at which Elizabeth draws upon decidedly different discourses to explain her practices and preferences. These are subtle variations, not marked by a conscious attempt to differentiate herself from her mother but perhaps pointing to generational changes in perspective. It is also possible that she draws upon her religious background to a greater extent than her mother, although this will be explored more fully in the final section of this paper.

Discourses and Media

Both Jean and Elizabeth speak of television as a bad habit. Elizabeth does not consider talking on the telephone, which she spends much more time doing, a bad habit, nor does her mother. Of course, there is much less stigma attached to phone than television use, so this is not surprising. Perhaps because of her former heavy watching of television during the lonely days of her marriage, Jean describes it as the worst habit anyone could have:

> There are other things to life. I think watching tv hour by hour by hour, you're wasting your life. You're wasting your day away...I think tv makes us vegetables, big time.

When I asked if she then thought television should be educational, she responded enthusiastically that she would be perfectly content with Discovery and the two public broadcasting channels, although she

admitted that she likes certain entertainment programs on television, such as the ice skating championships. Later, Elizabeth echoed similar concerns regarding television's ability to consume too much of one's time. When I asked her how television viewing compared with other bad habits, Elizabeth replied,

> It's not as bad as smoking (her mother is a smoker). Worse than nail-biting, I guess. But you can really get into the shows and watch them every day, that's really not a big deal if you're watching a couple shows, but if you're watching three hours at a time, it's like, there's more stuff to do.

Elizabeth, who demonstrates a great respect for her mother, does not differ from her mother's views on television, although she herself seems less interested in watching television (and thus less in danger of developing a time-wasting habit of tv viewing) than her mother once seemed to be.

While her mother dislikes television's ability to encourage continued viewing, Elizabeth identifies a different possible influence. After a conversation about a violent tv show, I asked: "Do you think that there are some teenagers who are negatively affected by media, and others aren't?" While I expected her to talk about some teens who might be more passive due to their viewing, or who might be more influenced by television to participate in violence, she responded:

> Yeah, there probably - there is. A lot of my - some of my friends think very negatively about Mexicans 'cause they hear so many bad things about Mexicans. Or they don't like this one [musical] artist just because something happened that, what they did or something. I don't know, I think they should read into it, see what's really going on, and why they did that or whatever.

Interestingly, a different discourse emerges here, one not present in conversations with Elizabeth's mother and one not directly prompted by our previous conversation. Elizabeth demonstrates a concern for the

ability of media (television in particular) to inform stereotypes which, in turn, result in what Elizabeth feels is the unfair passing of judgment on certain artists.

A muted discourse of tolerance emerged again in a later conversation with Elizabeth on the subject of a school in the Denver area which had received widespread publicity for the controversy surrounding the changing of its mascot from the "Redskins" to the "Reds." She noted that her father had attended that school and was really angry about the name change. He had threatened to go to a game again just to cheer for the "RedSKINS," she said. She related this story matter-of-factly, reserving judgment both about the mascot and her father's views, in contrast to her father's vocal opinions.

Discourses: Media as Distraction

It is of course common in U.S. discourse to describe the media - and television in particular - as "bad." However, people differ in how they explain why this is the case. Perhaps the most frequent explanation is that television is bad because it encourages immorality through the display of sex and violence.[2] A second frequent explanation is that television is bad because it represents certain groups unfairly, or fails to represent them at all, thus causing or at least reinforcing stereotypes and prejudices common in the culture.[3] Elizabeth and her mother, however, draw upon a third relatively common discursive strategy in explaining the failings of television: it takes time away from other, more productive activities.

[2] This viewpoint has also been the assumption underlying many "effects" studies of the media and children, such as in the studies by Bandura (1977) and Berkowitz (1964).
[3] This viewpoint has also encouraged research into media and young people, although less frequently.

This emerges in their conversation noted above as well as in talking about limits on television watching. While the Farleys have a stated rule limiting television watching to one hour a day, it is not enforced due to Jean's absence. It is also not enforced when Jean is home in the evenings, as she reads while John watches television in the living room. As she says, one hour a day "is ideally [the rule]. That means that when I'm not here, I'm at work, they need to pick and choose and use their time wisely. And if [watching television all day] is what they choose to do, there's nothing I can do about it." Elizabeth states that she "tries" to watch only an hour a day as her mother has requested.

Because Jean believes that television watching is not a productive use of time, she explains their family viewing habits with expressions of guilt or self-deprecating humor which imply that she feels she is not as responsible a parent as she should be. This was in spite of the fact that it seemed that she was frequently in the same room with Elizabeth and John during their viewing, thus "chaperoning" the viewing (which is the recommended strategy for parents concerned primarily with media content as opposed to those concerned with time).

There are no rules about television content, although Jean expresses her wish that Elizabeth would not watch adult soap operas and John would not watch *The Simpsons*. While Elizabeth agrees that she should not watch them and doesn't except on summer afternoons, John is more nonchalant about viewing programs of which his mother disapproves. They do not subscribe to cable tv, primarily for cost reasons.

Media as Educational and Therapeutic

Perhaps because she is interested in demonstrating herself as a media-savvy and responsible parent, Jean advocates educational television in her conversations with me. Elizabeth does not echo these interests at all and gives no indication of watching public or educational television on a regular basis. Similarly, Jean's explanation of her online use begins with an echo of the "educational" discourse employed earlier to describe her television preferences:

> Jean: See, I'm trying to go back to school, so I did use it for a little research on scholarships. And then when I got done with that then, yes, I too was using the Internet. The chat rooms and stuff. It was fun.
> Lynn: Oh, yeah, they're great.
> Jean: Stayed up one Saturday night from one in the morning 'til 7:30 the next morning.
> Lynn: What chat room were you in, can I ask?
> Jean: Ah, it was the After Hours. It was just one particular room, there was a lot of people there, and - you really get to know each other. Amazing the regulars on there. They're on there every day. 'Don't you go to sleep? Do you have a life?'

Note that what began as a description of a pleasurable activity ended with Jean passing judgment on the others participating in the chat room, perhaps to echo what she believes is the more acceptable view of the practice. In a subsequent interview, the educational use of the media emerged again, as did the therapeutic use of media to establish relationships. Jean noted:

> Jean: There's a special room that I really like to get into and that's the women's room, and they just talk about everyting from their kids to, to - divorce, sex, whatever. No raunchiness going on in *that* room.
> Lynn: There is quite a bit of raunchiness going on in other places, isn't there?
> Jean: Mmm-hmm. Yeah, if you feel like a little bit of raunchiness, any night after 10 o'clock, go into the After Hours chat room. But I like the Internet not only for that, but like when I was doing research on scholarships, research on just about anything. It's nice to know that

there's research here that you can access a lot faster than getting into the car and driving to the library.
Lynn: So when you were doing the chat things with the women, is that similar to the kinds of conversations you have with your friends?
Jean: Well, with my friends it's more intimate stuff.
Lynn: So the chat thing was appealing because -
Jean: Something to do. Anything to avoid housework! (laughs)

Like many others in the U.S., Jean is uncomfortable admitting that she enjoys media use as a pleasurable activity. Speaking to me as a relative stranger at the time, it is not surprising that she quickly turns the conversation from her experiences in the "raunchier" rooms of the Internet back to the safer turf of educational uses of the media.

Like her mother, Elizabeth mentions several ways in which the media serve the purpose of drawing her closer to others and strengthening relationships. While Elizabeth's emphasis is on her relationship to friends, however, Jean more frequently speaks of the "self-help" books she reads and the chat groups she has participated in as sources of support, although she, like Elizabeth, also used the phone to be in contact with friends after the divorce. Elizabeth's approach to the media thus seems arguably more integrated or embedded in her life, while Jean tends to see the various media as a possible disruption to other aspects of her life. This may be a generational difference, likely attributable to the differing roles of parent and teen.

Inxterestingly, the Farley's discursive strategy shares some common ground with the Frankfurt school and other neo-Marxist critiques of culture as distracting people from discovering the true nature of their oppression. Given the history of Marxist thought in the U.S., it is not surprising that the Farleys are not overly conscious of their oppression, although Elizabeth is aware of the financial strain resulting from her

parent's divorce and is able to relate this to her own possibly curtailed opportunities for college as a result. Yet their views are not antithetical to a Marxist position, although they take a distinctive twist in their particular context. For example, Elizabeth feels that television-watching time is better spent in building relationships in school and at church, which might be identified as a rather 'gendered' strategy which combats isolation and builds important networks of support for the present and future. She also spends time doing homework, which has a direct relationship to her ability to go on to college and possibly claim financial success beyond that which her father and mother attained (although, as noted earlier, her lackluster academic career may undermine this goal). Meanwhile, Jean also champions education as a means toward self-improvement, which she pursues through college courses, self-help books, and a support group. Thus, in their talk of television both Elizabeth and Jean in some ways embrace the Utopian vision of a better life which drives much of critical theory. Yet rather than addressing these concerns in the arena of public and/or political life, they are seeking to attain their "Utopia" through gendered and culture-specific strategies which are individualistic rather than collective, relying upon education rather than perhaps more radical political change. Because of the importance of relationships for both, it is also possible to argue that these strategies are primarily psychotherapeutic. However, in so doing I want to note that too often, women's relational strategies are dismissed as in some way secondary. Most critical theorists argue for the importance of solidarity among those who are oppressed, and thus it is possible that these relationships, like others throughout the culture, have the same degree of potential for politicization as any others -

and perhaps more, due to the encouragement for action found in women's groups (See, e.g., Griffiths, 1995).

Discourses on Gender

As noted earlier, Elizabeth is encouraged by her parents in her accommodative and passive approaches to others. This is also evidenced in her reluctance to confront her friend who has joined a gang, choosing instead to screen her calls until she "gets the message." Other ways in which Elizabeth speaks her gender include her discussions of her friends and their relationships, such as in her referral to the use of teen magazines to give advice to friends, and her discussion of "cute boys." Each of these, as well as the idealized female standard she constructs and claims as her own in her 'Net persona, speak to the naturalized discourses of female expectations in the U.S.. Heterosexuality is another discourse in which Elizabeth is embedded, which is expressed in her discomfort with photos of the female celebrities, the case of mistaken gender identity on the Net, and the rejection of the gay celebrity Keanu Reeves.

The Media and Religious Discourse

Elizabeth's narratives demonstrate the extent to which she understands religion as related to organized religion and religious *institutions*, and specifically, to her church. When asked if she considers herself "spiritual," for example, she replies:

> Elizabeth: Um, not really. I think of myself as religious, because if somebody's talking about their religion, I'll usually jump in and say, 'Well, at my church we do this.' But I don't know about spiritual. I haven't really gotten really, really into my religion yet. I haven't really found the spiritual part quite yet.

> Lynn: Is that something that interests you?
> Elizabeth: Yeah. I really like getting into the church thing. It helped me a lot through the divorce, 'cause half the people in the youth group have been through a divorce before, so I got a lot of moral support from them, and the youth leader was in a divorce, so I kinda got it from her perspective, too, so that really helped.

Thus "spiritual" is something separate from and beyond the religious institution. What has been particularly meaningful for her has been the institutional and therapeutic elements of the religious community, rather than a more individualistic aspect of "spirituality" such as in Bellah and his colleague's (1985) "Sheilaism" and Roof's (1993) description of the new-age "Molly."

When asked about whether or not she ever talked about religion with the boys she "dated" on the Net, she also responded within an institutional discourse of religion:

> I mentioned that I was getting confirmed, and I told him about, like, how much of a pain it was to go through the three years instead of only the two years, and he's like, 'Yeah I can understand that, that happened to me, too." So we kind of relate to church sometimes.

As noted earlier, Elizabeth is much more involved in her church than her mother or brother. I had therefore expected this connection to emerge with respect to either issues of morality or tolerance, as these tend to represent the two approaches to media taken by differing "sides" of the "restructuring" religious environment.[4] Yet I found elements of both approaches woven into her conversations about media, although there was little connection between the two. Religion was closely related to morality, and seemingly only negatively related to tolerance. When asked about whether or not she felt that religion influenced her media choices, for example, she replied:

[4] An extended discussionof religious discourse in the U.S. occurs in chapter seven.

> Not really. Sometimes if there's something that's religious on TV, that was like based on something that happened in some religion, sometimes I'll check it out. But I won't stop watching a show because it's like, something that's immoral or that wouldn't go by in our religion.

Religion, like other issues in the media, is filtered through a larger moral scheme of right and wrong. While she acknowledges the expectation that the portrayal of religion in media should be evaluated, she prefers to allow the entertainment value of the programs to dictate her viewing choices, such as in the case of *The Simpsons:*

> There's been, *The Simpsons* with the baptizing in it [Baby Maggie is taken away from Homer by neighbors to be baptized]. Homer didn't want his children to be baptized, which, I thought that one was pretty funny, but - I thought it was kind of pushing it, too.

Thus while her talk about religion in general had led to conversations about religious institutions, her talk about religion in the media equated religion with moral choices. I found this particularly interesting in that the moral dimensions of her religious institutional life and of her personal religious beliefs did not emerge at any other time in our conversations.

The most interesting aspect of Elizabeth's narratives on religion, however, emerged in conversations about the television shows *The X-Files, Inside Edition,* and *Touched by an Angel.* Each of these, she states, are programs about "beliefs" rather than about "religion." She explains:

> Elizabeth: *The X-Files,* it's kinda, not really what you believe *religiously,* but what you believe just, what you believe in, like ghosts and stuff. If you really believe there's ghosts, or extraterrestrials. Same thing with Inside Edition...
> Lynn: Do you ever talk about media representations of religion with your friends? Like we were just talking about, like on *The X-Files,* or *The Simpsons?*
> Elizabeth: Sometimes it'll come up, like we'll ask each other if we've

seen that *X-Files* thing or something, then we'll talk about it...

Elizabeth then went on to describe a paranormal experience one of her friends has had, stating that she "can't rule it out" as a possibly authentic experience. On the other hand, however, she stated, she is very skeptical about the portrayals seen on *The X-Files.*

"Beliefs" are more "open" than religion is, according to Elizabeth. Although Elizabeth identifies herself as "religious," by which she means affiliated with organized religion, she feels that her views on religion are more similar to *Touched by an Angel* than to *Soul Man,* the 1997-1998 prime time program featuring Dan Akroyd as an Episcopal priest and single father. She states:

> *Touched by an Angel,* it's kind of open on who God is, 'cause it can kinda be whatever your religion is, but with *Soul Man,* it's kinda set with, the guy's a priest, or minister, or whatever, so it's pretty much set that he's either Catholic or Christian. But with *Touched by an Angel,* I mean angels, you don't even have to believe in God. So you can just believe that there are angels....*[Touched by an Angel]* can be any religion, and it doesn't have to be religion, it can just be a belief.

From the program's use of the word "priest" and the fact that the Akroyd character's supervisor was a bishop, Elizabeth surmised that the church depicted was Roman Catholic - despite the fact that Akroyd's character is a single parent. This illustrates her limited knowledge of religious belief systems within the U.S., a limitation to be expected among a generation less familiar with institutional religion than perhaps any previous one in the country's history. I thought it was very interesting, and revealing, that Elizabeth would find *Touched by an Angel,* a program I will describe more fully in chapter seven as rooted in Judeo-Christian beliefs, to be "open" to various belief systems. Because there is no mention of the

institutional church, it is not a "religious" program but one of "beliefs." "Beliefs" are optional and less "set," according to Elizabeth, a view that seems to articulate the "seeker" mentality Roof (1993) describes. It is interesting to me that Elizabeth, an avid adherent and advocate of organized institutional religion, finds the portrayal of angels in *Touched by an Angel* compelling. While she recognizes that they are not consistent with the Christian understanding of angels, she has no apparent problem appreciating and holding both beliefs simultaneously.

Another interesting comparison emerged between *Touched by an Angel* and *Soul Man* in Elizabeth's narrative. One of the reasons she liked the former more than the latter was that she described the didactic *Touched by an Angel* as "not really preaching, like *Soul Man* is." While some episodes of *Soul Man* included a moralistic epithet delivered from the pulpit, every episode of *Touched by an Angel* includes the delivery of a two-minute monologue in which an angel speaks to the central character about God's love. Thus even though the angel's conversation included language that was equally if not more explicit in its Christian focus, and the program's tone of delivery was more serious than that of the sitcom *Soul Man*, it was not seen as "preaching." "Preaching," according to Elizabeth, is limited to that which occurs behind the pulpit, just as "Christianity" is limited to that which occurs in the context of overt religious symbols such as a church building and a central character in a clerical collar.

I believe that Elizabeth's narratives illustrate the ways in which Christian belief systems are naturalized within U.S. society to such an extent that in the case of Elizabeth's talk, there seem to be only two

categories: that of religion, which is the equivalent of the Protestant Christian religion, and that of "beliefs," or those outside the Christian norm which might also be popularly classified as "superstitions." In chapter seven, I discuss this hegemonic view of Protestant Christianity in greater detail. It was largely because of my intriguing conversations with Elizabeth on these topics, however, that I was able to conceptualize the focus groups with both teens and adults, and how they might illuminate differing discursive strategies relevant to contemporary public representations of religion and spirituality in the U.S.

Discourses of Tolerance and Multiculturalism

While morality is subsumed into the norm of Christian religion, tolerance, conversely, seems to have little connection with religion in Elizabeth's narratives. This differs from the frequent assumption that mainline religions such as that to which Elizabeth belongs are more oriented toward justice issues (Hoge, Johnson, & Luidens, 1994). As is consistent with a general discourse of tolerance in the U.S., Elizabeth is concerned about the injustice that she hears and sees in the media. When discussing a rap singer's lyrics, for example, she notes:

> He's referred to black people as "niggers." I don't remember what the song's called. And I was like, I don't like that word. That was my first reaction, I was like, "I don't like that word. I'm not gonna listen to that song." [Her best friend] was like, "Well, I don't like that word either, but I like the beat to the song. I don't really listen to the words." "I sure hope you don't listen to the words."

While Elizabeth has embraced tolerance, at least in her narrative self-presentations, as a "good," her views seem to be at odds with those expressed by her parents, friends, and church members - although she does

not seem to realize this. When I asked her if her mother would agree that "nigger" is a bad word, for instance, she at first agrees, but then adds a caveat and changes the subject:

> I think she would, probably. She doesn't - she's never really had good experiences with black people [she then describes a single bad experience her mother had with a group of African American girls when in grade school]...But none of my friends have a big problem with black people...

Of course, the fact that none of her friends "have a big problem with black people" also suggests, first, that none are black, and second, that none of them particularly value or argue for tolerance, either. In fact, she notes that they "sometimes get kinda snotty" about the fact that members of the same racial/ethnic groups, such as Asians and African Americans, choose to sit together at lunch. Elizabeth herself uses linguistic "codes" in her references to race and class, such as her earlier referral to Shaw Heights - a code for the rough and largely Mexican-American city high school attended by the gang member who had called her. Bringing the conversation again to her parents, I asked whether or not she has ever had to correct her mother or father in their use of discriminatory words. She replies:

> Um, I had to, once, with my dad's girlfriend [who is a recent immigrant from a Scandanavian country]. She made a comment that America was only for Americans, and that all black people should go back to Africa. And I decided that I didn't like that very much, and I said, 'Well, you know what? You're not from this country, so maybe you should just watch what you're saying."...And I probably would have been more aggressive, but my dad was giving me the look, "If you get too aggressive, you're going to have to go to your room." So I just spoke my mind and I just said, "don't say that. I don't like it." And she's never said anything else.

The fact that Elizabeth's father was angered by her statement suggests that

he is less interested in teaching his daughter tolerance than in teaching her to be a "proper" non-aggressive female. Yet it is also important to note that the passion of Elizabeth's outburst was probably at least as much due to her dislike for her father's girlfriend as for her concern for tolerant language.

Her church members, as well, seem less interested in tolerance, and less able to empathize with those understood to be "others," than she is:

> Lynn: What would your church think about, Mexicans, for example, or blacks?
> Elizabeth: Um, I think the older people have more of a problem with it. But our neighborhood, our church is kind of in a neighborhood where there's mostly Mexicans. And we now do have a black girl going to churh. And on Sunday, she got up in front of the whole congregation and read a reading. And I thought that was pretty courageous, 'cause if I was the only white person in a black church, I would not have the guts to get up there and do something like that.

Probing the tolerance of the church community further, I asked whether or not the church ever discussed the Mexicans in the neighborhood:

> Um, they [the Mexicans] kinda keep to themselves. But my youth leader has mentioned, if more people from the neighborhood do start coming to the church, would we have a problem with it? And none of the kids had a problem with it. So I really don't think, if we had a problem, if there was somebody that came into the church, I really don't think that there would be any discrimination, except for maybe with the older people. They're kinda like, 'This is a white church. Only white people will go to this church.' But those older people are moving on, and we're getting more and more younger people, so. I think we're [younger people] more tolerant. Because the older generation, it was kind of shown to them that black people or Mexicans were lesser than us because they were from a different country. Or because they're a different color, or they speak a different language. And I see that a lot in church. Usually they try not to show it, because of the new girl. You know, she's new and we don't want to drive her away, because she's got a big family and if we drive her away, we're gonna drive the whole family away. So, they're being pretty nice about it, but I think deep down they have a slight problem with it. But with us, there's, I've got a few friends that are black, I've got a few friends

> that are Mexican, I've got a few friends that are Asian. And it just kinda ranges, it doesn't matter what they look like, to me, and to most of my friends, they don't care. Unless they've like had an especially bad experience with those kind of people, which I haven't heard any of.

In this comment, we can see that Elizabeth's definition of discrimination is much more narrow and confrontational than the more subtle forms of racial injustice experienced every day by non-white groups, or as she says, "those kind of people." While attempting to portray her church as tolerant, then, she has revealed just how hostile the environment is to non-whites. We also see the limits of the tolerance as practiced by her friends, and by Elizabeth herself.

Thoughts on the Researcher/Participant Relationship

On the drive home from Elizabeth's confirmation at her church, my husband, who had accompanied me to the event, joked, "it doesn't seem to matter whether you're a researcher now or not; you always end up being a mentor to young people." He was referring to my decade-plus years of experience leading church groups of junior and senior high students, as well as my experience as a junior high and high school tutor through various civic organizations.

Over the course of two years, I formally interviewed Elizabeth once with her family, four times individually, twice specifically about the peer-led discussion groups, and twice within the context of peer-led discussion groups. I also met with her many more times first to help me plan and lead the pilot and then the first peer-led discussion group, and later on social occasions. These latter meetings were not tape recorded, although I took notes following each session. While her mother was initially

skeptical of the interviewing process and of me as an interviewer, she has come to trust me and sees me as in an informal mentoring relationship with Elizabeth. This is probably an apt description of how our relationship has evolved, although I argue that there is still a formality to the relationship that is not present in other, more explicitly "mentoring" relationships I have had. This is the case for several reasons. First, I never would have met Elizabeth outside of the research project. Second, while we enjoy getting together for social occasions, I have also asked for her feedback on my research goals, and have distinguished "formal" interviews from more social events by bringing along my tape recorder. We also frequently e-mail each other, and our relationship through this medium has ranged from me helping her with homework assignments, to me soliciting her opinion about new television programs or tentative conclusions in my research, to an exchange of cyberjokes. I think the relationship is beneficial for both of us because I find Elizabeth to be an interesting, mature, and reflective young person who I like.

Her involvement in her church and her media use are in some ways much like my own were at her age, and this makes me conscious of my own assumptions going into this research project and of how vastly different my own experience was from the current teenager's. Like most young people, I believe Elizabeth enjoys being seen as an expert and appreciates the fact that I listen intently to what she has to say. It is an asymmetrical relationship, typical of research or therapeutic relationships in that one person does more questioning while the other does more answering. Yet like all researcher/researched relationships, there are points of tension for me, particularly around the issue of politicization.

As a researcher informed by a critical framework, I can analyze the issues of inequality based upon class and gender which inform Elizabeth's experience, as well as her unchallenged assumptions about race. Yet what is my role in educating her about what I see? I think that this is particularly challenging in the research of teen girls, for I know that because Elizabeth values our relationship, she would not want to say things that offend me or which might cause me to question her values. Thus I reveal my own beliefs at the peril of learning hers. I have reconciled this to myself by holding back on self-revelation.

While I am skeptical about much of the political approaches in cultural studies as discussed in chapters two and four, I do hope that this project, which seeks to expose the discursive strategies surrounding media and the relationship of these to other aspects of individual and cultural identity, might contribute some understanding of how and why such strategies "work" hegemonically and hence practically.

Chapter Six

Parents, Teens and the Media

This chapter serves as a further contextualization of the experience of identity-construction among U.S. adolescents. As I noted earlier, one of the benefits of studying adolescents is that, as they most frequently share a household (or households) with their parents and siblings, it is possible to gain insights into the larger familial and other structural processes shaping their experiences by interviewing other members of their family. As I am interested in exploring their processes of *religious* identity-construction in particular, I wanted to probe the discourses parents employ when discussing issues of religion, the media, and their adolescent children. There are several discursive strategies which may be said to be related to larger public discourses on these topics, and thus I will begin with an exploration of these themes. I will then explore what the parents felt were the most salient issues in these areas. As these issues, I believe, in many ways frame and shape the teens' discourses on media and religion, this chapter will conclude with a foundation for comparison with the next chapter, which explores in greater depth some of the similarities and differences found in the conversations with teens on these same topics.

Parents and the Discourses of Media, Religion, and Adolescents

As noted in chapter three, I organized three parent discussion groups toward the end of my research so as to maximize my ability to

utilize their expertise on life with teenagers, and to see if I could affirm or challenge some of my own growing understandings of teen life based on my interviews and reading. As noted in greater detail in chapter three, one of the groups was organized for me by Judy and Tom Ortiz, and consisted of their friends and family members, all of whom are Mexican American conservative Roman Catholics from lower middle or working class backgrounds. Another group was organized by Lester and Ellen Gray, and consisted of their friends, all of whom had been members of the same Presbyterian (U.S.A.) church for which Lester is the former pastor. The members of this group are all moderate-to-liberal upper middle class caucasians. The third group was organized by Cathy Swenson and consisted of her friends and associates, all of whom participate in a parents' support group for divorced or widowed persons. Six of the members of this group are caucasian and one is part Native American. All might be considered lower middle or working class except for one who has a higher income and education level. Four identified themselves with Protestant churches although only one actively participates. One member was observant Roman Catholic, and two were not associated with organized religion.

In organizing the parent groups, I was centrally interested in four research questions: (1) What do parents want to teach their teens about religion? (2) What do they want to teach them about the media? (3) How do they use the media to teach about religion, if they do? And finally (4), as I am interested in generational differences, I also wanted to gather their reflections on how their relationships with their teens differed from the relationships they had with their own parents during their teenage years.

It is this latter category that I believe provides the most intriguing insights concerning how contemporary teenage life in the U.S. may be different from earlier generations, and thus also how the media play a role in the current identity-construction process.

There are several discursive limits to the current topic that emerged in the conversations with the parents, and these relate to the public understandings of religion, the media, and adolescence outlined in earlier chapters. For example, many of parents I had interviewed tended to equate religion with values, as I will describe more fully in a moment. They also articulated the discourse of the media's negative influence discussed earlier - even though their actual practices, some of their beliefs, and even several more subtle descriptions of interactions with teens around media issues did not always support this discourse. Because of the concern for the possible behavioral effects of the media on their teenage children, however, they tended to approach the questions surrounding the media practices of their teens as ethical issues. These discursive limits understandably framed my conversations with them, as it seemingly frames the conversations they have with their teens on these topics.

Religion was a particularly difficult issue to discuss, which is perhaps not surprising due to the popular taboo against such discussions and the trend toward privatization which places religion increasingly in the realm of private and personal choice, and the varied definitions of religion embraced in the U.S. One parent even suggested that perhaps I should use the term "spirituality" rather than religion, as that was what he believed I *really* meant by my use of the term. Indeed, although employing both the terms "religion" and "spirituality" was helpful with

some parents, I found that there were others who held differing approaches to religion for whom this was problematic or simply confusing.

The Discourses of Religion

As Roof (1993) notes, people in the U.S. approach religion along several dimensions, including a tendency to see religion as defined by (and often rejected because of) its association with the institutions of organized religion, as an important aspect of racial/ethnic or community identity, as a way to approach life's existential questions, or as a part of an overall search for the meaning of life. While I found examples of each of these approaches in my interviews, the conversations about teenagers, religious identity and media often highlighted an understanding of religion as that which provides a moral foundation for their teens. To some extent this affirms Hoge, Johnson and Luidens' (1994) assertion that Americans believe that an important function all religions share is the upholding of a "moral code." Thus while I often tried to guide the conversation regarding issues of religious (or spiritual) identity toward issues of power, contestation, or negotiation with other aspects of identity, or toward existentialist questions of "meaning," it was often the moral issues that framed and shaped the parents' conversations. The following excerpt from the Ortiz's group discussion illustrates this common discursive approach to religion as a moral system:

> Lynn: Are there some things about religion that you try to teach your kids? And how do you try to do that?
> Connie: You know, what's right, and what's wrong, what's good and what's bad. You know, that there's just certain things that you do, you don't want to hurt people, you don't want to hurt yourself because

you're a temple of God.
Judy: We always try to teach our children that it doesn't matter if we know what you've done, whether it's good or bad or whatever, because God always knows everything.
Cheryl: God always knows.
Judy: - and he's the one that matters, he always knows what you're doing.
Connie: It's between you and him.
Judy: So, when you go out there, whatever you're doing in life, remember, consult God first.

In order to pursue my interest in religious identity-construction, therefore, it became necessary to explore how and in what ways *values* were discussed by the parents in relation to issues of both media and religion.

The discussions on media and values tended to echo the discourse of negative influence which was discussed earlier. When asked about whether or not the media have an influence on their teens, for instance, all but one of the parents in the groups emphatically agreed that it does. Judy and Tom Ortiz, the Roman Catholic Mexican American couple who hosted one of the groups, expressed the belief that one of their primary parental tasks is to provide moral teaching to their two teenage children. They believe that their religion assists them in this, and they see the popular media as rather directly undermining their efforts. Note how they explain one "values" conversation with their 14-year-old son:

Tom: I mean, we're all media people, we were brought up with television, and believe it or not, we're influenced by television since we were children....[...] The reason I say this is, one time Jordan brought something home and he goes, 'here.' And I go, 'what is this?' And he said, 'Oh, we did it during 7th hour.' And it was sent home from a public school and it had to do with religion. And some of the things - I said, 'did you read this?' He said, 'yeah.' And some of the things were off the wall! One of the things was -
Judy: It was about living together without being married. Is that okay with you, is it not okay? It was okay with him. And in these situations, what would you do if your children had no food on the

table? The only way to get money is by selling drugs. Would you go and sell the drugs? Well, he said yes, and we were floored! We said, 'Wait a minute! We've been teaching religion in this house since the day you were born, you haven't missed mass unless you were sick! Where are these ideas - ?' 'Well, mom, realistically, if your kids are starving and that's the only way you could get food, you've got to feed them!' That was his reasoning. And I said, 'Nobody in America will ever starve because you can fall over and be at a church, and they'll open their doors and you get fed!' The process is to *think* about it, not to react to the situation but to *think.*
Tom: You see, so even though I think we've been teaching him, and doing this - [the media have an influence].

Judy and Tom are appalled that their son Jordan expresses views that differ from what they see as values rooted in their Catholic faith, in this case that living together without marriage and selling drugs are always wrong no matter what the circumstances. They see their son's regular participation in mass as a primary means by which he was to have learned those values. It is worth noting, too, that Judy points to the ablity to rely on the charity of the church as a means of resisting the need to sell drugs for money, illustrating again the central role of the Catholic church in supporting what they perceive are its (and their) values. Yet for the current project, perhaps what is most interesting is the rather direct link Judy and Tom make between their son's failure to express their preferred values and the influence of the media. In this stance, neither the environments of peers or of school are considered, nor is the possibility that the church (or the parents) have somehow failed to convincingly socialize Jordan into what the parents perceive as desirable values rooted in Catholicism.

Religion was also discussed among the Gray's group. Like Judy and Tom's son Jordan, Tanya and Jack's son Andy - who incidentally is the same age as Jordan - has also expressed some differences of opinion,

although in his case it was more specifically related to religious identity than to a moral code:

> Lynn: Tanya, do you want to comment on your family's approach to religion, in terms of what you're trying to teach them?
> Tanya: Well, I guess at this point Andy has rejected organized religion. He would not voluntarily go to church with me at this point.
> Lynn: How old is he?
> Tanya: He's 14. And he assures me that he has a spiritual nature, but it's not in church.
> Lynn: So, how do you talk to him -
> Tanya: Well, I hope he has a good foundation, if he ever wants to return to that, but then, you know, I think I observe his spiritual nature. So, just encourage him to explore. I'm certainly not going to force anything.
> Lynn: Does he look at other religions?
> Tanya: Not at this time. And that's something I think we'd be willing for him to do.

For Tanya, while religion may be the foundation of a "moral code," she believes that her son is old enough to make up his own mind and therefore does not force him to attend services with her. While he does not seem interested in exploring other religions, it is perhaps the fact that he has a "spiritual nature" that affirms his moral code. It is also possible that having a "spiritual nature" is, according to his parents, another acceptable form of religion that may or may not be at all related to morality.

I commented in chapter four on the parents' belief that teens need to make up their own minds about religion. As I noted there, two of the mothers in Judy and Tom's group stated that their teens were encouraged to explore other religions. Although they hoped, like Tanya, that their children would one day "return to the fold," they also believed that forcing them to attend religious services could engender a rejection of religious life altogether. The desire for information on other religions

was encouraged, among both those with a religious affiliation and those without it. In the Gray's group Don, who grew up without religion but now has an "intense interest in it," is married to Julie, who grew up and seks to instill in her children strong ties to organized religion. They agree that "open discussion" is the central characteristic of their approach to religion. Similarly, Melanie (also in the Gray's group) notes:

> Melanie: I've gone to many different churches, and I try to explain my beliefs to them...And I always send them out for a reference. Maggie started going out with a boy that was Mormon, and so she wanted to know more about the Mormon religion, so she said, 'can I go to church with him?' And I said, 'Absolutely.' And this other friend, her best friend, said, 'oh my God! You wouldn't let her go to a Mormon church?!' And I said, 'one of my best friends is Mormon,' and I sent her to Kelly's, and Kelly sat down with her and explained the basis of their religion. And so that was real helpful. But that tends to be how we talk about it - it's kind of a reactive conversation.

Melanie indicates that religion is not something that emerges in her conversations with her teens very often, although she seeks to affirm their search for information when it does. The same approach is articulated by Justine, a woman in Cathy Swenson's group who is an ex-Catholic and now has no religious affiliation:

> Justine: I never stopped [my daughter], if she wanted to go to church, you know, she could. That was totally up to her, but - I think religious-wise, she's kind of decided that organized religion isn't for her, right now, either.

Contrary to some of the other parents, Justine draws a firm line between religion and morality, as she wants to encourage the latter but not necessarily the former. Yet she, too, affirms the approach of encouraging her daughter to search out religion and to decide upon it for herself. Lester, who as I noted earlier is a pastor in a liberal Protestant denomination, commented on this ethic of encouraging an "openness" to

exploring different approaches to religion in his tradition and among his own teens:

> Lester: In our tradition, there's an encouragement to question and to search for truth and understanding, which is integral to the system. Some people don't follow that, but it's integral to it. And so when I talk about church to our kids, I talk about those kinds of values in our way of looking at things. Look at the Bible with a critical eye, try to understand...It was important for me in raising my kids to say, 'this is how we do it.' Because at some point in their life, they all have to know, 'what does my group do? What does my family say?' And that's an identity issue. Every one of us needs that. And then if you pursue the strong questioning, searching for truth, and your own understanding, then it really does take you beyond the denomination. And I think it takes you into the kind of realm that [fellow group member] Don's entered into, where you start thinking, 'what are the similarities? How can I synthesize this?' And that's an exciting dimension, too, but then churches and denominations tend to want to reach out and grab you by the ass and say, 'come back here! What are you doing? You're thinking too much!' Because the system has to keep perpetuating itself. So that's the love/hate relationship that I have, that's the struggle growing in the denomination that wants to propel you out, and then at the same time, pull you back.

Lester's comment helpfully broadens the analysis from the realm of individual parental choices concerning approaches to religion to the fact that this orientation is central to a tension in the moderate-to-liberal wings of American mainline religions today. As young people are affirmed in their ability to make their own choices concerning religious identity and religious affiliation, it would stand to reason that the distinctiveness of different religions would continue to decline. The 'search' results in something like "cafeteria Catholicism," Bibby's (1987) term for the practice, not limited to Catholics, of picking and choosing the most personally meaningful aspects of the larger faith system while discarding those elements that are less desirable. Andrea, the 18-year-old daughter of Lester Gray, certainly articulates this approach as a result of the

search she has been encouraged to pursue in the realm of religion. She notes that she appreciates being able to talk with her parents about religion, but states, as did a number of other teens in my study, that it is in the discussions with her peers that she learns about and makes her own decisions regarding religion. "We kind of make up our own religion," she says.

Some of the parents in Cathy Swenson's group equated the search for religion with a search for meaning among young people. This group speculates that perhaps religion is fulfilling a different cultural role for today's teens than it did in the past, and rooted its appeal in the need for both security and an alternative to the violence and chaos common to their everyday lives:

> Joe: I think they're looking for something to hang on to...
> Sandy: We have one of the largest teen groups in our church, it's huge. And they have a mass just for the teens. And that's on Sunday night at 6, and they have a band for the music. And the parents were asked that if they come they sit way in the back. But I mean, that place was rocking, and I was absolutely shocked, and it used to be like 15 years ago, the dorks went to this kind of stuff.
> Cathy: Right.
> Sandy: But now it's just, a whole mixture of kids, and the church is packed. So I do think kids want this and I think they're bombarded with so much garbage, grossness - . [...]
> Cathy: Right, it's in the music, it's on the television, it's in all the movies, and I think the kids too are kinda - . With my daughter Brandy, she loves to go to church, and I think she just has that void. And for some reason, maybe it was the way I grew up, my parents' disinterest in religion, my yearning for it, we went to a church that was 'hell and damnation,' you're bad if you do this, you're bad if you do that,' and I became kinda jaded and a little bit, not real sure of religion because that was the perception in my mind, all they want to do is pass the collection plate, take your money and tell you how horrible you are. Where Brandy hasn't been jaded that way, and she's so open to just drinking in whatever she can get out of church, and like Joe, it gives them hope, and something to put their faith in, that there is some purpose, that there is a reason for things, and that there's someone

watching over you all the time. I think that gives her a lot of security, especially as a child of a single parent. She needs all the security she can get a hold of. [...]
Donna: Maybe other things are coming in and taking the place of some of the things that we were used to. Since schools do seem kind of disconected, they're getting bigger and bigger, and kids are working, and they're not using the school site as much as the family site, maybe they're going to the other stuff. You know? Maybe it's okay to go to Girl Scouts, maybe it's okay to go to church.

The parents in this group present the teens' various sources of support and identification as relativized or flattened to some extent. The once-regnant hierarchy of the "acceptable" versus the "dorky" activities in which teens may participate is less determinative, these parents note. Teens echo this sentiment in commenting on the sense that the boundaries of "cliques" seem much more fungible in their own day than they did in their parents'. Thus just as the choices within a religious system are open for negotiation, so the choices of activities for affiliation seem less restrictive and limiting, and are approached with the same measure of pragmatism. Of course, there are social and structural forces that shape and subtly inform the choices teens can and do make; none of these caucasian teens "choose" to attend predominantly African American or Mexican American churches or social clubs nor is organized religion appealing to the all adolescents. Still the perception among both teens and their parents is that the choices available to teens today are relative; each activity may be acceptable if it is found to fulfill the desired functions of providing security, meaning, social networks, activities, or a combination of all of these. It is thus not so much that organized religion either is or isn't a viable avenue for meaning, but that it is on equal footing with several other options. It is to be judged for its effectiveness in meeting

these needs, according to a number of the parents and teens with whom I discussed this issue.

Judy and Tom, of course, see their religion as something other than simply one of many options with which to fulfill needs. Yet in their group, their tendency to equate religion with morality led to several extended discussions of the need for morality, and thus even their conversation was framed by this functional approach to religion to a great extent.

Dealing with the Media

The emphasis on morality also continually led the conversation in Judy and Tom's group back to issues of competing influences. Evoking the "negative influence" discourse, for example, Diane and Rob, members of Judy and Tom's group, expressed their concern about the media's influence on their teenage daughter with reference to her practices of rebellion:

> Diane: Tabitha's been influenced by what the teenagers do on those [talk] shows.
> Lynn: In what ways?
> Rob: 'Cause she's a single child, and she has to act -
> Diane: No, she has to interact with the tv because she has no siblings, so she has to interact with the girls on these talk shows -
> Rob: And then she tests it on us. We're the testers.
> Lynn: What do you mean?
> Rob: Oh, little things, like talking back.
> Diane: Running away. She threatens that.
> Rob: She just keeps going and going, to see how far she can get.

Like Judy and Tom, Diane and Rob seem to believe that their daughter's undesirable behaviors may be attributed directly to her media practices. Diane also indicates that perhaps 16-year-old Tabitha's status as an only

child makes her more vulnerable to identification with those in the media, as she does not have siblings with which to "interact."

To counteract what they perceive as the media's powerful and influential role in their children's lives, parents take differing approaches. Some parents in the Gray's group, all of whom were upper middle class caucasians affiliated with the Presbyterian (U.S.A.) church, note that they sought resources from their church and from parachurch organizations to provide "positive" impressions. Melanie, for instance, employed *Focus on the Family* videotapes during her children's younger years, while Lester and Ellen note the positive influence of popular Christian crossover singer Amy Grant. Yet this approach sometimes backfires when the youngsters enter their teen years, as is evident in Judy and Tom's experience:

> Judy: Couple of years ago, there was this production company that wanted to send us movies, family-type movies, and I could not get my kids to watch - they just laughed at me. We ordered the movies, and -
> Tom: The Butterscotch Gang -
> Judy: Butterscotch Gang, and - they just looked at me, and they said, 'you're kidding!' Misty thought it was the funniest thing!
> Cheryl: And what is it about?
> Judy: The Butterscotch Gang, they all had messages, all clean, those people sold me on the idea that, you know, you gotta implement this in your home. And I said, 'that's right, we do, we have to, there's gotta be a message in everything we watch.' So they were good, clean, fun things that the kids did a bad thing and then they worked through the problem, and my daughter just could not believe I would even pay good money for this. 'You really don't expect us to watch this?!' She just laughed at us! I don't know, their mentality's not there for that, or the structure of the family is so much that - . (shrugs)

Judy and Tom do not watch television or movies with their teenagers because they say that their tastes differ dramatically from their teens.' Surprisingly, they actually exercise very little limitation on the viewing

practices of their 14- and 16-year-old. The Ortizs note that there is a television in every room of their house, so the parents have limited ability to supervise viewing. Thus the teens usually make their selections independently, a practice that seemed fairly common among the families with teens I had interviewed. Other parents note that rather than watch programming with their teens which they are often not able to do, they prefer to use the teen's request to view certain television programs or movies as opportunities to express their views, particularly on sexuality and violence. In the Gray's group, Julie, for instance, opted for censoring one popular teen program:

> Julie: [14-year-old] Abbey and I sat and watched *Dawson's Creek* because it comes after *Buffy [the Vampire Slayer,* on the WB Network]. And after the first two episodes, she and I had this long talk, and I felt it was inappropriate for her to watch, because - one of the moms was having an affair, the one student had an affair with a teacher over the summer...

Rather than outright censorship, Donna, in Cathy Swenson's group, chooses to encourage her teenage son to exercise good judgment and to understand the effect of his choices as a consumer on the overall mediated environment:

> Donna: I'm pretty dismayed by the crap I see out there right now, and when [17-year-old Bill] chooses a movie, I've said time and time again - . You know, he'll try to defend what he's going to see, and I'll just say, 'Bill, remember if you're not against them, then you're basically with them. And it's your decision, you're 17, but whenever you spend your money to rent that video or to buy that book, or to buy that CD, whatever you're spending that money on, you're encouraging more of it, so just think about that.' And usually, I'd have to say he argues with me and does what he wants to do anyway (laughs). But I know he thinks about it, and there's been a couple times when he's come back and said, 'Mom, you're right, I've got to think about that.' Or he'll see a movie, and say, 'Mom, I know you'd approve of that movie,' so I know it's sinking in.

Of course, the encouragement of such judgment may not be possible with teens in their younger years, such as Julie's two children.

A few parents indicate that watching television programming or movies with their teens is one of the primary ways in which they are able to engage them in conversations about values. In Cathy Swenson's group, both Cathy and Sandy note that the news often sparks conversations with their 16-year-old teenage daughters, particularly with reference to tragic stories or those stories dealing with teenagers. In the Gray's group, Melanie and Jack complained that they have felt compelled to discuss sexual issues with their teens "like oral sex, and what is adultery, and all of those issues because of our president and the media," as Melanie says. As Melanie's teens are 12 and 13 and Jack's is 14, they believe that their teens should be protected from these issues. Ellen, whose teens are 18 and 11, also laments having to deal with the sexuality in the media, although she notes that it has become a positive opportunity for her to converse with her girls. In discussing the program *Dawson's Creek,* for instance, she says:

> Ellen: It raises so much sexual - . We spend time, 'well, what does *that* mean?' I'm like, [shoot, I have to get into this!] Andrea says, 'I *think* I know what that is.'...I've had quite a lot of talk, on one level, with Andrea, at 18, which was sort of a different conversation than Sally at 11. And I appreciate it, because it's an opportunity for me to hear their reaction, and then I can give some feedback. And it became positive. And so every Tuesday night, we sit down and watch *Buffy* and *Dawson's Creek.*

These comments led other mothers in this group to reflect on their own use of the media in discussions of values. Julie, for instance, rearticulated her censorship practice as an opportunity for discussion, and Tanya, who seemed less certain of her own practices, also explained her stance *vis a vis*

the popular and controversial animated program, *South Park:*

> Julie: Well, and I've questioned my censorship, too, I mean, I tell her, 'explain to me why this is a thing you'd really like to watch, now the discussion's open.' And I'll say, 'it's just that I felt that several of the issues - ' that she wasn't giving me any feedback on the issues either, she was just internalizing what was happening on the show. She did say, 'look, Mom, it's not something I'm gonna go out and do.' 'Well, I appreciate that, but I don't feel like you have to be exposed to that.'
> Lynn: Tanya, you were going to say something.
> Tanya: I was just wondering, I walked into the room, and Andy had on the cartoon South Park, which he'd gotten a copy of, and I was, my eyes were [bulging]. I said, 'where did you get this, Andy? What is this?' And he said, 'a friend.' Of course, we're the only people who don't have cable tv.
> Julie: We don't.
> Tanya: So I think I said, 'this is a good reason not to have cable.'
> Julie: Have you heard of it before?
> Jack: I've heard of it on the radio.
> Tanya: No, well, I have since, but I was - [shocked]. I couldn't believe what I was seeing. So I would definitely restrict that, but I guess he can go over to his friends'. But he sure knows how I feel about it.

We see in this and earlier conversations three differing approaches to expressing opinions concerning the media: viewing programs with teens with the intent of guiding the viewing experience through discussion; censorship with a discussion explaining the parents' views; and latent permission for the teen to continue viewing but with the knowledge that the parent disapproves or at least wants to encourage an exercise of judgment on the part of the teens. It is worth noting that in these excerpts as well as in the vast majority of the family interviews I conducted, it was the *mothers* who articulated the family's stance with regard to media and its use or limitations.

While decisions around media selection and practices serve as an important time in which parents can communicate to their teens what they consider to be important values, there are at least two other times

when these conversations occur. First, values-related discussions emerge for many parents and teens when a rule is broken or when an action warranting punishment occurs. Tom articulates this in his group's discussion:

> Lynn: How do you teach [16-year-old Misty] what's important, or what to value?
> Tom: I think, what I do is - . That's why after I spanked her, I talked to her, to let her know. 'Cause the thing I remember about my dad was there'd be times where he would give me some pearls of wisdom, little things, and I remember those. But it was when he was talking to me, not when he was hitting me.

After noting the importance of discussing the reason for punishment, Tom then referred back to the earlier conversation regarding Jordan's school handout on moral issues, noting that he would rather have values discussions when they emerge spontaneously than when a boundary has been crossed. Lester and Ellen highlighted these spontaneous conversations on values in their group discussion, noting that such conversations emerge in relation to activities their teens are engaged in, such as retreats they attend with their peers. While other parents note that they attempt to share time with their teens either around the dinner table or before the teens depart for school, Lester and Ellen are the only parents who seemed to foreground such family discussions with a great degree of intentionality; as Lester says, "Ellen's really good about sitting us down and saying, 'let's talk about this.'" Perhaps it is fitting, therefore, that the place most parents reported having similar conversations was when they were already seated with their teens - in the car. After hearing this from several parents during individual interviews, I asked each of the discussion groups about it. Here are some excerpts on the subject from Cathy Swenson's group:

Sandy: I think you're correct about the car. That's the one thing I missed when she turned 16 in January, she started driving herself to school, and that used to be - it gave us time to say, 'what's going on,' you know. And I really miss that in the morning.
[...]
Donna: [The first time Bill drove on his own], I got on the phone and cried to my sister. I said, 'he just left, and I am totally out of control now! I cannot do anything!' But I think the driving thing is - . 'Cause up until my divorce I lived out of state, and we spent a lot of time, long, extended periods of time in the car, and I miss that road time. 'Cause when you're in the car, you have a captive audience. No phones, you know. And even - since we're talking about media - even when you've got the radio on you have to be listening to the same station. Because he can't be in his bedroom listening to whatever while I'm in here - so I'll be able to ask, 'why do you listen to that crap?' And he'll go, 'Mom, it's not crap, listen to this.' Or 'that's kind of a cool song, Mom.'
[...]
Cathy: Well, talking about the car, that's been a good time for us, and I think a lot of it is the music, the radio. And so many of the songs are remakes from the 60s and 70s. And I start singing along and Brandy looks at me and goes, 'Mom, you're listening to my music?' And, 'no, it's a remake,' and she goes, 'no, it's not!' And I say, 'your generation is so unimaginative they can't come up with their own songs!' But then we start talking about, 'oh Brandy, I remember when this song was on and I was hanging out with my friends and I was wearing my POW bracelet and my Levi's and my hiking boots, and I was so cool, 'and she's looking at me like, 'oh, God.' But those are good conversations, you know, and they lead to other things, where she'll say, 'my friends said this,' and we'll have some real good conversations.'

The other two groups affirmed the importance of car time, as well. As Tanya notes: "It seems that if something's bothering one of the kids, that's a time when they know they have my attention, and no competition." This comment points to the fact that while we often think of *parents* as pursuing the opportunities for such discussions, the busy lives of today's parents also require that *teens* seek such moments for reflection, discussion, and attention to their concerns as well.

As noted earlier, only one of the many parents interviewed argued

against the negative effects of the media. Cheryl, Judy Ortiz's younger sister, continually combatted the "direct effects" comments of her sister and others in the discussion group, arguing for the influence of the broader cultural environment on teens:

> Cheryl: Yeah, see that's what I'm saying, it's [the negative influence] out there, so no matter what we censor on our tv or on our whatever, it's not really gonna affect - . I think it's how you teach your kids. [...] I let my kids watch like, *Colors,* and *Boys in the Hood,* all that stuff, we watch it together. Because regardless, I never even seen [those movies] and I knew about gangs and stuff before I even seen the movie, because the way the kids have been into that scene already. So, it's gonna be out there regardless. [...] I don't let it [media] influence my kids. I tell 'em, 'is that what *you* really believe? 'How do you know that's *really* happening in California?' 'Well, they say there's this - ' 'But how do *you* know?' Be from the show-me state. Always question it.

While the other parents in her and other discussion groups were not forthcoming in their reports of television and movie viewing practices, Cheryl enthusiastically discussed her enjoyment of various movies and even the frequently-denigrated genre of talk shows, sometimes weaving in comments on discussions that resulted with her son and nephews as a result of these programs. While her discussions were perhaps less intentional than those of other parents, she indirectly used the media to generate discussions about meaning and values.

Although all but Cheryl affirmed the negative influence of media and expressed concerns about it, there were thus varying methods of approaching the issue. These included consuming media together either with or without discussion, seeking "positive" substitutes for the "negative" popular media, outright censorship, the more subtle drawing of distinctions between parents' and teens' preferences, the affirmation of the teens' ability to exercise good judgment, and intentional discussions

encouraged in relation to mediated texts. While I had expected that the parents who expressed the greatest concern regarding the media's potential negative influence would be the ones to exercise the greatest restrictions, I did not find this to be the case consistently, nor did I find that those who were least convinced of media's effects were also the least involved in their teens' decisionmaking regarding media. Clearly, it is an unresolved issue, and in part the ambivalence results from the fact that the parents themselves grew up embedded in a mediated environment.

Parents as Teens, and Teens Today

Most of the parents believe that their lives are busier, and that their teens' lives more full of distractions, than was the case when they themselves were teenagers. One surprising learning for me was that in each of the groups, the parents spoke nostalgically of the time when the entire family used to sit down together in front of the television set. Now, parents note, teens are more likely to be in their rooms alone, listening to music, talking on the telephone, or perhaps watching television or a movie. Surprisingly, there was less consensus on whether or not media content was "cleaner" then than now, however. Judy believes it is much worse now: "Gol, look what's on tv, and remember when we were little? They used to, you know, to say even a cuss word, just was never heard of!" Yet others approached the issue with more relativism as they reflected on the restrictions their parents had placed on their own viewing practices as teens, and on their parents' approach to the media more generally. Lester, for instance, believes that his parents viewed the media with a greater degree of mystery than his own generation does:

> Lester: I think they came in an era where the media was sort of mysterious, the radio was a mysterious kind of thing, and the television was a mysterious kind of experience, and it was not just entertainment for them, but the equivalent for them of the book, with respect to the source, the meaning. So, things that came across in the media were things that were to be believed, and to be - . And so they were more open to it. We're the network generation, which sees the foolishness of it, too.

Other parents, while affirming Lester's insight that the media were new to their own parents and thus were perceived as powerful conveyors of information by them, do not see their parents as more "open" to the media than they themselves are. Instead, they believe that their parents' naivete concerning the media led to the belief that the media were a great danger to their children. When today's parents were teens, then, they adamantly fought against this belief. As Melanie says in the Gray's group:

> A comment that my mother and mother-in-law say constantly, 'I don't think those kids should watch television.' I gotta tell you, a kid who doesn't watch television is very unusual in this day and age.

Melanie, Julie, Ellen, Tanya and Jim all note that their parents had resisted (but finally relented to) their desires to watch The Beatles on the Ed Sullivan show and thought Elvis was "vulgar," for instance. Julie comments on the attempts at persuasion that finally resulted in the compromise situation concerning the Beatles in her household: they would watch the program as a family, and would turn the tv set on just before the show and turn it off immediately after. "And I remember the click on the tv, when the tv went on," she recalled, "That was impacting!" Judy and her sisters recall that they had been forbidden from watching the soap opera *Peyton Place,* even though they knew that their mother sometimes snuck inside to watch it in secret without them. Cathy fondly remembers watching evening television programs with her family,

although noting that their father maintained strict control over viewing based on his own preferences and the family never discussed anything viewed together. These experiences of being on the receiving end of the discourse of negative effects, I believe, resulted in the increased ambivalence I observed in issues surrounding the media. While the parents still articulate the discourse of negative influence, their practices were actually varied and support a much more nuanced understanding of the media-adolescent relationship.

The parents in Cathy's group equated the teens' comparably autonomous media consumption with a general increase in independence among today's teenagers. They note, in fact, that the very character of their relationships with their teens is quite different from the relationships they once had with their own parents on this point:

> Justine: I don't know, I'm seeing a pattern here. Where my parents had - I don't know if authority is the right word, but they made the rules.
> Joe: They had control.
> Cathy: Yeah, control. They said, 'this is the tv show we're gonna watch, the telephone is for our use, you can make a quick call and say, 'hey, Susie, I'll meet you at the corner -.' But you know, my daughter ges on and talks to one person for four hours. And I have to go to my daughter and say, 'Brandy, get off the phone, I need to make a call.' That's real interesting, I never thought of that, but it's almost a reverse of when I was growing up. I would have to ask permission to use the phone, and then it was quick, on and off.

This group also noted that they believed that they participate in more activities with their teens than their parents had with them. As Justine recalls:

> You know, I remember when I was a teenager, I never did anything with my mother. I mean, what was I gonna *do* with her, she didn't - . (laughter) But you know, my daughter and I, we go to the show a lot, and it doesn't phase her to go places with me, or tell people that she went to the show with us.

Sandy notes that while her parents always attended the events in which she and her siblings took part, her mother would not have participated in the activity with her, such as the skiing and tubing adventures Sandy has shared with her own daughter. On the other hand, Judy and other members of her group affirm that her teens think that "it's not real cool to be seen at the movies with your parents." This difference in joint activities and perception may, in part, be related to the fact that the members of Cathy's group were single parents. In a sense, then, they had the added difficulty of setting limits while also maintaining close relationships that in many ways flattened the once-hierarchical relationships they had known with their parents. As Connie, a single parent in Judy and Tom's group, notes:

> Connie: sometimes you can almost be too friendly with your kids. And then when you need the discipline, you have to kinda be, well, 'wait a minute, I'm really your mom, remember?' I have the say. You can't push things too far, I'm still your mother, and you need to have respect for me, and this is the rule, and what I say goes this way.

The struggle over discipline and the desire to balance this with a meaningful relationship with their teens was apparent in all of the groups to some extent. When I asked whether or not they believed that they had the same level of influence on their teens that their parents once had on them, Judy and Tom's group emphatically agreed that they did not, and they equated this lack of influence with this changed nature of discipline:

> Cheryl: I think with Nardo I do, but my other two? 'Cause our parents had *that look.* You knew you were in trouble if -
> Connie: My dad was so mean. Oh, he was the scariest man in the world. But he's a different person now.
> Tom: He still has that scary look, though.
> Rob: But I think they had to be mean because -
> Cheryl: But I think the rules have changed, too -
> Rob: Oh, the rules have definitely changed -

Cheryl: And I think the kids use that. I don't think my dad was mean, I just think we were scared to mess up, 'cause that's how we were brought up.
Judy: I think the rules and the whole lifestyle growing up was a lot different. It's really relaxed. It's an anything goes attitude now, where that wasn't the attitude when we were growing up. You know, when your parents spoke, you listened, and when there were people in the house, you were seen, not heard.
Connie: I think my relationship is different, it's like, I'm more involved with my kids than I think - . My parents were my parents. And that was it. You didn't have a friendship or anything with them. You didn't know their business, you didn't ask them their business. And they didn't come into your life, or try to be your friend, or try to know anything.
Judy: But there were expectations, you were to go to school, do what you were supposed to, and the parent was working and doing their thing, and I think we're probably different, not better, but different because we knew that it was important to be involved with your children, know what they're doing at school, be there for PTA meetings.

Single parents such as Connie and the members of Cathy's group saw themselves as conflicted about discipline and also more interested in having a relationship with their teens than they believed their parents were. Even members of the two-parent households affirm the challenges of discipline, however. As Rob, a member of Judy and Tom's group, says:

Rob: I read an article one time that said you should let your kids talk back, because as they grow into adulthood, they will not be afraid of authority, to express themselves, and I was raised to where you were seen and not heard... So I think, I want her to respect us as parents, but how do you do that with letting her talk back? You know? So there's a fine line. So I'm battling with that. I just want her to be strong, that's all.

Tom also emphasizes the ways in which practices of discipline have changed, and thus how his approach differs from his father's which then also has implications for the issue of authority:

Tom: I never spanked them the way my dad did to me. They were scared, and I would, you know, but I think the only time I ever spanked

Misty was when she was what, 15?
Judy: And then he had to explain to her why he was going to give a spanking, for an hour.
Tom: I needed to. [...]
Judy: I think the difference between our parents and the way we are, you know, our parents only spoke once, they only said, 'this is not right, don't do it because the consequences will catch up.' And they meant it. And today, it's like, 'if you continue to do that, you're gonna get a spanking, one more time -'
Cheryl: Yeah, 'no' was 'no' back then.
Judy: So many threats and there's no follow through to that....

Members of Cathy's group also talked about the power of "the belt" to instill fear in them as young people. Cathy articulates the conflicted feelings already mentioned:

> Cathy: We were afraid! And I think there's a good and a bad side to that. I was afraid of my mom, I didn't have a loving relationship with her at all, where I have a very loving relationship with my girls now. And it's totally different than with my mom. I don't know if there's that much level of respect. I don't know where you draw the line between respect and fear. I think they respect me, I don't think they fear me, but sometimes fear is a motivator, so I don't know. But I'm much more comfortable with this relationship, I would never want a relationship with my daughters [like the one] I had with my mother.

Another difference between their experience and that of their teens arose in relation to the larger community. When they were teens, today's parents could walk or ride their bike to their friends' houses, they stated, whereas now parents note the need for teens to call their friends to plan their outings before arriving at someone's door. This was related, again, to the issue of discipline. In Judy and Tom's group, Cheryl noted, for example: "When we were young, the teachers would always get us in trouble, and the lady down the block would get us in trouble, and now it's always on the parents. See back then, it wasn't, it was all the parents." In Cathy's group, they discussed this shift as a way of reemphasizing the

autonomy of young people through what they believed was an increased use of the car.

Changes in the structure of the family were also observed as a difference shaping the experiences of the teens today. This was of course clear in the single parent families, but the parents' opinions concerning the impact of these changes was not at all consistent across all of the parents. A number of the single parents pointedly spoke against the stereotype which defines families such as theirs as "dysfunctional," for instance, and I observed that this stereotype was indeed inappropriate for most of these households. Yet Jack, one of the parents in the Gray's group who has never been divorced or widowed, made the most pointed statement I'd heard with regard to these changes:

> Jack: I'll tell you the biggest difference in the neighborhood that I grew up in, and the one I live in. In the neighborhood I grew up in, there were *normal* families.
> Lester: Normal?
> Jack: That's a bad word, but there were families where there was a husband, wife, and kids, whatever they call that kind of family. Now in my neighborhood, I have all these *other* kinds of families. I have some families like that, but we have lots of these, what do you call it? *Blended* families. Boy, we got a bunch of those, probably half our neighborhood is blended families. And we have all kinds of strange things that we didn't used to have.

As I have earlier advocated a somewhat reflexive approach to methodology, I need to comment on my own response to this remark as I hope it will shed insight into the ways the generations following the "baby boomers" approach such issues of family. My own adolescent experience was shaped in a two-parent, first-marriage family much like the "normal" families Jack describes. However, I was so offended by his comment that I did not realize until after I had transcribed the group discussion that my

approach to Jack shifted dramatically after that point in the conversation. I became impatient with his tendency to answer questions indirectly and even interrupted him a few times to basically seek a more direct answer to my question, call on others, or to change the subject. As I look at the other transcripts I realize that my response to Jack was, thankfully, fairly uncharacteristic of the way I handled other situations in which I disagreed with someone. Perhaps this is due to the fact that while I had expected (and indeed did hear) many racist remarks from (primarily) caucasians of various socio-economic backgrounds, I was caught off guard by this judgment of what I grew up considering to be *normal* families: those like my own, but also those that are headed by a single parent, or by two parents previously married to others, or by persons not married, as well as those of differing religions and racial/ethnic heritages. In fact, my own definition of *normal* families now also extends to the gay, lesbian and bisexual couples that are raising children. Although I am reluctant to identify myself as a member of "Generation X," I can affirm anecdotally that I know many persons my age who also would have bristled at Jack's comment (or at comments passing judgment on marriages of mixed races or religions, which I heard from other members of the parents' generation). I can therefore surmise that the situation of varied families has become even more common and acceptable to today's teens - with the caveat that I recognize that teens live within a strong heterosexual norm that might make them less comfortable with gay, lesbian or bisexual parents. However, I believe that most of the teens I met could become open to interactions with gay, lesbian or bisexual parents if they had a friend who came from such a family, due to the ethic of tolerance they are

encouraged to embrace.

The Discourses of Tolerance and Diversity

This approach to tolerance for persons of differing backgrounds is another significant difference between the environment of today's teens and that of their parents. While parents expressed varying degrees of tolerance themselves, they all generally agreed that their teens were more accepting and that this was a mark of progress. In Cathy's group, for example, Justine and Donna note this difference:

> Justine: I think kids are more tolerant of each other. I see it, anyway.
> Sandy: They want to be.
> Donna: I think it's a huge push in the schools. I mean, you go to elementary to high schools on up, everywhere there's 'teach tolerance,' 'Celebrate diversity.'

Lester and Ellen's group related this tolerance to television viewing preferences. When I mentioned a *New York Times* article (Hass, 1998) on the proclivity for caucasian teens to watch certain programs with all-African American casts and for African American teens to watch certain programs with all-white casts, Julie noted:

> Julie: I'm not sure they even take note of that. You brought up the fact that it's all Negros in one evening. I'm not sure they know, other than the media advertising that night, I don't think they even realize that -
> Lester: Consciously.
> Julie: Yeah. From show to show. They look at it and see people. They either like the show, it's stupid, or it's not.
> Lynn: Is that different from what you'd see?
> Julie: Sure, I notice when somebody's black. You know?

This same issue was also raised in Judy and Tom's group, although rather than an erasure of color they saw a tolerance for differences:

> Rob: I think that Latino kids are raised different than black kids, and black kids are raised different than white kids, and when you put 'em

all together, they seem not to worry about that.
Lynn: They seem not to worry about what?
Rob: About how they're raised. As parents, we worry about it more.
Cheryl: Yeah, I see that a lot around my neighborhood, 'cause we're mixed [racially]. And it's not a big deal 'til somebody comes into our neighborhood and makes a big deal out of it. The kids don't really, you know, see it.
Judy: I think that's what this generation has up on us.
Cheryl: Yeah.
Judy: Is that color is not an issue. And it might have been more of an issue for us, and even a worse issue for our parents.
Cheryl: 'Cause our neighborhood does well. We're majority black, but there's Chicanos, some white, and some Vietnamese. I think the worst thing I've ever seen happen is when this little boy came from Texas, and where's he's from there is not mixing, there's only black. And he's like, 'how could you be playing with a Mexican?' My kids were like, 'it's not a big thing, you know? What's he talking about?'

Although the parents primarily talked about tolerance in terms of racial/ethnic diversity, when asked about religion they echoed tolerance here, as well:

Judy: My children do not distinguish that, 'okay, I can't hang out with my Jewish or Muslim friends, we're just too different.' They're best friends. They leave their - they explore their issues.

The tolerance for persons of differing religions seemed to dovetail with the encouragement they gave their own children in their religious search. They all affirmed my comment that their teens live in an environment that is more diverse in terms of racial/ethnic and religious backgrounds than the environments in which they themselves grew up. Yet the different approaches the parents took to these issues was perhaps most apparent when they discussed the excerpt of the television program I had selected for the teen peer-led discussion groups.

How Adults Viewed Touched by an Angel

As I noted in chapter three and will discuss more fully in the following chapter, I had chosen an episode of the popular television program *Touched by an Angel* as an intervention for the discussion of issues of religion, media, and diversity among teens. I showed this episode to each of my case-study teens who in turn showed it to their discussion groups. Thus, I felt that it would be interesting to show an excerpt of the program to the parents' groups to learn how they, too, responded to the program - and how they may have expected their own teens to respond to it.

Rather than review the entire 60 minute program, however, in the parents' groups I discussed the premise of the program and the specific episode to be shown, and then played the final 10 minutes of the episode (a full discussion of the episode will be offered in the next chapter). This concluding segment included some elements common to all *Touched by an Angel* episodes, as one of the lead angels, Monica, reveals herself as an angel to the troubled African American youth and gives him encouragement to make a life-changing decision to care for others rather than continue in his self-destructive ways. The segment also includes an interaction between the youth's mother (a single parent) and another woman concerning single parenthood, and then concludes with the tearful and changed youth's return, signaling his positive decision to care for both the disabled girl he had become involved with as a volunteer, and, by extension, for others such as his mother.

I had several reasons and reservations regarding the use of this particular program, as I will discuss more fully in the following chapter.

However, I also had certain expectations regarding the reception of the program by the parents' groups. I had believed that the parents groups would, in general, like the program more than the teens did. This was not uniformly the case, however; in fact it seemed that whether or not a person liked the program was definitely related to gender among the parents' groups (the women liked it far more than the men), while in the single-gendered teens' groups there was a less clear division between male and female viewers. Instead, in these groups it was the persons who saw some similarity between their own religious beliefs and those portrayed in the show that liked it more than those who saw no such similarity. Of course, it would have been interesting to see if the male teens would have been more critical of the program in a mixed-gender situation; I suspect that this would have been the case.

Based on how the preferences among the teen groups related to their approach to religion, I had anticipated that of the three parental groups, Cathy's group of single parents would find the greatest resonance with the program based on both its apeal to conservative morals and its positive message about single parenting. I had expected the Gray's group of moderate-to-liberal Presbyterians to be the most critical of its sentimental religion, and the Ortiz's group to be most critical of its representations of race and class. Instead, the moralistic theme of the program held the most appeal for the Ortiz group, but it was the group of Presbyterians who were so moved by the program that they asked for tissues to be passed around. I was struck by how uncritical they were of the program's central themes, particularly when compared with the other groups. While the women in this group unanimously agreed that it was

"touching,"some women in the single parents' group felt it was "mushy" and "unrealistic," for instance.

In fact, the criticism that the program was "unrealistic" was one of the most common critiques mentioned by both the parents and the teens. Yet after hearing elaborations on this comment, it became evident that the measures of "realism" differed within the various groups. Both members of Cathy Swenson's and Judy and Tom's group, for example, note that it would be "unrealistic" for a troubled youth with a history of stealing cars to return this one and completely change his life, and Lester Gray expressed a similar sentiment in his group. When watching the program, in fact, a few members of Cathy's group commented incredulously on the opening visual images: " He steals cars from a dealership?!" In contrast, the Gray's group indicated that it was what they called the "cliched" God language, the lack of a reflection of the "complexity" they experienced in their own lives, and the fact that their teens did not know anyone like the troubled youth, that made the realism problematic. As Don comments:

> Don: Well, see, I think our kids are fairly well sheltered. I don't think they've run into that particular characteristic of a teenager who really has no expectations of what his life is going to be like. I think all of my kids' friends expect to go to college, and grow up and have jobs, as opposed to the ghetto attitude of 'what difference does it make 'cause I'm never gonna get anywhere.' I just don't know how they'd respond. Abbey (14) would not really understand where this comes from. Denny (16) would think it's all too hokey.

The fact that the members of the Gray's group saw themselves as more distant from the story than the members of the other group is illustrated when Lester brought up the subject again, this time in relation to the complaint raised by several of the parents concerning the sexualized issues of early 1998 newscasts:

Julie: [commenting on her reasons for censoring the popular teen program *Dawson's Creek]* : The one student had an affair with a teacher over the summer.
Melanie: Yeah, but that happens in the news! I mean, we heard about that teacher -
Jack: We have our President and Governor! [referring to the well-publicized accusations of sexual impropriety for both]
Melanie: I'm sorry, they just ran that story about that teacher, now she's pregnant again!
Lester: You're kidding!
Melanie: Six months pregnant!... But this is the stuff that comes on every day, and we have to discuss issues like oral sex, and what is adultery, and all of those issues because of our president and the media.
Lester: Yeah. And those are ethical issues that are much different than sitting in a car and saying, 'am I gonna steal this car?' They're much more complex, it seems like.

Lester is referring, I believe, to his conviction that the problems of the teen boy are solved simplistically in this program. Yet perhaps rather than only being less "complex," the problems associated with car theft are also less frequently associated with the trials of middle-class life. In contrast, the experiences reflected in the program seemed to resonate somewhat more with members of the other groups. Cheryl, in Judy and Tom's group, identifies with the program's protagonist, for example:

Cheryl: I liked how it gave the message.
Lynn: What did you see as its message?
Cheryl: That he was saying that, nobody should love him. that - how can I say this? - you know, he was thinking that nobody cared, because maybe his mom was saying, 'no, he couldn't do this, or he couldn't do that.' He thought she was being strict so that's why he would go steal the cars and everything. And the surroundings, probably, where he grew up, it's like, you know, you're a man, you're supposed to do this and that, and a lot of peer pressure. He didn't really give it a chance. Like at the end he said, he started missing people. And he probably realized then that his mom was being right, and his friends were doing wrong. [...] I think when you get older, you realize, because, I know I was the black sheep of the family, and with my mom (starts to get choked up) I realize now what she did for me. And that makes it kinda hard. 'Cause then you start to say, 'she was there for you.' And when you have your kids, you see they tried the best. So that part of it got to

me. When I see people, when they get older, they realize that this, you know, she was right.

She interprets the show as affirming her mother's strict parenting practices, as well as her own. Given the ambiguity noted earlier concerning issues of authority, punishment, and the quest for meaningful relationships with their teens (which in the course of the discussion, occurred after the viewing and response to the program), perhaps it is not surprising that it was this aspect of the program that triggered an emotional response for her. It also seems appropriate, then, for Sandy to highlight the approach to single parenting she believes was embraced in the program:

Sandy: I thought the episode was good, I think as far as being a single parent - sometimes single parents, like that woman, get a bad rap that, if your kids don't turn out right, it's becaues you did something bad as a parent, and since you're a single parent, there's something dysfunctional about the home. And the most functional kids I know personally come from a one-parent home....I'm not saying all the time, everybody makes mistakes, but just because you're a single parent doesn't mean that if your kid doesn't turn out the best, it's because you messed up as a single parent. Some kids are just, their environment around them, and their peers, it's just too strong. But I think sometimes single parents go the extra mile because the kid doesn't have two parents.

One need only recall Jack's earlier comment on "normal" families to gain some insight into why this aspect of the program may not have resonated for him or other members of the Gray's group.

Several members of the Gray's group liked the fact that the program highlights important contemporary "issues," however, as Julie, a somewhat regular viewer, and Jim, who had not seen the program before, state:

Julie: In the past, *[Touched by an Angel]* has also brought in gang

violence and things like that. I've seen them bring in issues from today's world that I am glad to see them handling on this show. [...]
Lynn: Did you want to add something?
Jack: I think it's important for things like that to be on television, and I think that particular story is probably, something that needs to be talked more about. We do have, the black youths of America is almost an endangered species, with gang violence, and I don't know, that's a major, major problem in this country, and there are some good things being done, but it's a huge, huge problem.

The single parent group saw the benefits of the program on a less abstract level. Rather than "important issues," they liked the way the show portrayed "real-life situations," as Sandy notes: "I like *Touched by an Angel* because it shows real-life situations that, it seems like there's no way that things could ever be better, and they get turned around." When another parent in the group protested that expecting a troubled youth to change is "unrealistic," Donna responded:

Donna: Like Dick was saying, 'that's not reality, in reality a lot of kids would go off and [steal] the car.' Yeah, probably, maybe most, but I think maybe some of them wouldn't.
Joe: I like to think that maybe there's one kid who wouldn't do it.
Donna: Yeah, and I think that's the point of the show, that one of them didn't, and that's the hope you're talking about.

While Tanya, a member of the Gray's group, notes that "it is hard to comprehend that black youth have no hope," several members of Cathy's group seem to understand this lack of hope among teens only too well. In fact, they commented on the importance of holding out "hope" for teens and thus point to this aspect of the episode as a worthwhile possible "effect" of the program. As Joe and Jean note:

Joe: There's so much garbage on tv now that gives kids - I mean, kids are looking for a role model, or something to identify with, and there's so much stuff on tv that really takes them in the wrong direction, and this is more along the lines of giving the kids the sense that maybe there is some good left in the world. [...] The kids are goin', 'what have I

got to look forward to?' And these shows, they show that there's something. [...]
Jean: I think a show like this, it's kind of mushy, and not very realistic, but it's nice to know that there are shows out there that don't have the sex and the violence. It's nice to have shows like that once in a while that's just - fun.

Judy and Diane, from the Ortiz's group, each made similar comments:

Judy: We complain so much about what's on tv, but if we had more of that [shows like *Touched by an Angel*], maybe our kids would be tuned in with that. [...]
Diane: I try to look for the best in any kind of situation anyway, so it kinda went really good, and was a good ending. And whatever situations in the news or problems that happen during the day, I try to have that kind of hopeful outlook, you know, something good will come out of it.

Thus while members of the Gray's group may hope that such a program would acquaint their teens with important issues, it seems that the parents of teens who come from less privileged backgrounds hope that positive role models - whether they be in television or elsewhere in their lives - might help them deal with the similarly difficult issues in their own lives.

Another difference between the Gray's group and the other two groups has to do with the response to the God language of the episode, as noted earlier. Not everyone commented on this aspect of the program despite some prompting from me, again perhaps due to the sense that opinions on religion are to be private. Nevertheless, Sandy makes this comment which hints at the connections between the program and her own convictions:

Sandy: I'm religious, so I do believe in angels, and I do think there's always a reason, God always has a reason, even if it doesn't make sense at the time. You might say, 'why am I given this cross to carry?' But it makes sense later in your life.

Similarly, in the Ortiz's group, Connie also notes the importance of the God language in the show:

> Connie: What I got out of it was that she [the angel] was telling this kid that no matter how alone he felt and stuff, that God was always there, no matter how fast the car was, no matter how fast you go, God's gonna be right there with you. And I don't know what Steve [her son who previously watched the program with her] got out of it, but I always try to teach positive thinking, so I hope he got the same thing I did.

Yet in the Gray's group, the religious language contributed to the show's lack of realism or "hokiness," as Ellen and Lester both say:

> Ellen: The only thing that stood out to me as hokey were those phrases that are so cliche. 'Lost and he is found.' Things that in our frame of reference, we've heard so much, but someone who hasn't may react differently. [...]
> Lester: This [program] didn't quite do it for me, it wsa just a little too pat. I think, well, they tug on the emotions, but it's a little too, 'yeah, I've heard that. God loves you.' But would the kid really change? (laughs) He'd take off.

Thus while most of the parents agreed that youth's conversion experience - the premise of the program - was rather implausible, their responses to other elements of the story are what distinguish their reactions.

In both Judy and Tom's and Cathy's groups, there were elements of the program that affirmed their identity positions and thus created openings for the show to be made meaningful to them as the program articulated these positions. The program also articulated identifying positions for members of the Gray's group, allowing them to stand perhaps at a greater distance from the "issues" the program portrayed and thus reaffirming their positions within the status quo. The ability of the program to articulate such differing identification positions is in part what makes the program so popular. One specific aspect of this polysemy of the

text is its ability to articulate a variety of religious discourses, as well, which will be the topic of the next chapter.

Chapter Seven

Religion and the Media: Teens and the Public Discourses of Religion

The 1996-1997 television season ushered in an interesting era for those exploring the relationship between the media and religion. Even as CBS's*Touched by an Angel* became the second most-watched drama in the U.S. - regularly garnering an estimated viewing audience of twenty million - religion became a subtext for numerous other popular programs, as well. From *Touched* spin-offs and copycats such as *Promised Land* and *Early Edition*, to the clergy-centered *Nothing Sacred, Soul Man, Good News* and *7th Heaven,* to the religious questionings of main characters in such programs as *E.R., The X-Files,* and even *Seinfeld,* popular religiosity had found expression in the seemingly unlikely medium of prime-time television.[1]

Who is watching these programs, what do they mean for their audience, and what does their popularity reveal about the status of religious discourse in the U.S. today? Using the popular television program *Touched by an Angel* as an intervention, in this chapter I examine the interplay of religious thought and popular culture, relating the narratives of audience members to published narratives on the program, thereby attempting to describe certain aspects of the wider public discourses of religion. As noted earlier, I am using the term "discourse" here to describe the patterns of speaking of religion that are assumed to be "common sense" approaches to religion, in this case those accepted by

[1] Seven prime time television programs on four networks aired in the fall of 1997: *Touched by an Angel* and *Promised Land* on CBS; *Nothing Sacred, Soul Man* and *Teen Angel* on ABC; *Good News* on UPN; and *7th Heaven* on WB.

various groups and individuals within U.S. culture.

While previous studies of religious discourse have focused on the use of media by religious organizations attempting to influence public policy and journalists' interpretations of private religious experience in the public realm (Hoover, 1995; Hunter, 1991; Wuthnow, 1988), I am intererested in how individual experiences and expressions of religion are informed by and understood within larger public discourses of religion.[2] In this chapter, therefore, I will expand earlier discussions concerning the ways in which young people construct their religious identity, focusing on how they interact with the discursive themes on religion available in the public and mediated realm, and with their peers who often do not share their religious background or beliefs.

As noted earlier, the popular discourses of religion that frame the individual's notion of religiosity are not created by the media, nor are they even first expressed there. They are a part of the consensually accepted configuration of symbols, beliefs, and values that make up the totality of "popular" religion in U.S. society. In his definition of "popular religiosity," Lippy (1994) refers to such a "central zone" of religious symbols. He notes that while these symbols may be provided by and borrowed from formal religious traditions, "what is held in common may not receive systematic articulation or even rational justification by the religious elite of any one tradition, let alone by the ordinary people of a culture (p. 10)." My argument here is that the media, through their institutional position within society, have the power to articulate and provide legitimacy to certain ways of framing these symbols into

[2] For a more complete discussion of discourse and its relationship to power and identity-construction, see chapter two.

consensually-accepted and naturalized discourses.

Thus the media do not create popular religiosity, but might rather be understood, following Newcomb, as a public and cultural forum in which conflicting ideas of value are expressed and negotiated Newcomb & Hirsch, 1983). Echoing the consensually accepted "zone" of religious symbols is at least one important factor in achieving widespread success for popular television programs that deal with religion, therefore. Otherwise, such programs would appear unrealistic, offensive, or just silly and nonsensical, and could not garner the ratings necessary to survive. Those television programs that have some resonance with the audience's existing meaning-making systems, such as *Touched by an Angel*, therefore, point us toward these consensually-accepted discourses of public religion. These programs, in turn, inform and reinforce religious identity at the individual level, so that while each individual's religiosity may have unique features, there are also many shared patterns of symbols, values, and beliefs. An examination of the discourses articulated in *Touched by an Angel,* paired with the analysis of how people talk about the program - both with the interviewer and among their friends - therefore reveals something of these wider shared patterns of religion as it is popularly approached and articulated today.

While there are some important shared approaches to religion that are echoed in the program and in conversations about it, there are also at least three different discursive approaches to religion that appeal to different audience members of *Touched by an Angel* as revealed in conversations about the program. These also might be considered central discursive strategies of popular religiosity. I will refer to these as a

sentimental, a *revelatory*, and an *institutional* discourse. The individual viewers may identify with and employ any one or all three of these, as the text is polysemic and this particular program affirms all three positions.

This chapter therefore contributes to our understanding of "popular religiosity" by attempting to describe the discourses framing ways in which people talk about religion in the current U.S. context. It finds that each of the discourses affirmed by both the television audiences and the program itself are, not surprisingly, vaguely but primarily Christian. Thus even as Christianity is stripped of its institutional moorings, it continues to be conceived of as a "generic" religion in public discourse. The institutional and cultural authority of organized religion may be in decline, as theorists on both sides of the "secularization debate" have argued,[3] yet Christianity continues to hold a prominent and hegemonic position in U.S. culture, albeit in a changed and more diffuse form. It is this diffuse form that this research explores through an examination of the discursive stategies on religion.

Touched by an Angel : A Textual Analysis

To determine which media text I should use as a provocation for the teen peer-led discussion groups, I took copies of episodes from both *Touched by an Angel* and *Soul Man* to Elizabeth's home, where we

[3] The theory of secularization is outlined in Berger (1969), as was described in chapter two. Warner (1993) and Yamane (1996) each provide important analyses of the secularization "debate." Yet, as Tschannen (1994) argues, regardless of how they perceive the status of religion in the wider culture, scholars on both "sides" affirm the decline of the authority of religious institutions throughout culture.

watched both shows, as I described in chapter five.[4] Elizabeth preferred *Touched* over *Soul Man,* as she felt the latter was more geared toward an adult audience. Having expected that she would like the more humorous program, I was surprised at her choice. I was even more surprised when, serendipidously, Michael, another of my "case study" teens, named *Touched by an Angel* as his favorite program before I had told him of my interest in the show. Elizabeth has noted that since the discussion group, she has become a regular viewer of the program. While Jake, another "case study" teen, does not view it regularly, I was quite surprised by the degree to which he found the program enjoyable. Even Hasan, who I had chosen specifically because I knew his religious beliefs would be at odds with those portrayed in the program, suggested that he liked the episode after his second viewing of it.

Although the bulk of its audience is older than teens, *Touched by an Angel* grew in popularity among the 18-49 and younger age groups during the 1997-1998 season, and has somewhat regularly featured teen characters in key roles.[5] The introduction of a "teen angel" in the fall of 1997 marked an additional effort to appeal to the teenage audience. The high volume of fan mail and web page visits by teens also suggests that while the teen audience may be comparatively small, it is a program that holds meaning for some, particularly - as this chapter argues - for those who embrace one of the common discursive approaches to religion that are found throughout the wider culture and echoed in this program.

---

[4] *Soul Man* was an ABC program airing on Wednesday evenings after *Home Improvement.* A late introduction tothe 1996-1997 schedule, it is perhaps best described as *Home Improvement* in a clergy collar. It has been promoted by *Home Improvement*'s Tim Allen, who made a guest appearance on the show last spring. Dan Akroyd plays the clergyman and father.

[5] See details on its Nielsen ratings at http://funnymonkey.com/johndye.htm

To introduce the specific episode I selected, I would first like to review the central characters of the program. Tess (Della Reese), the head angel, is an archetypal/stereotypical black Earth Mother, overseeing the other angels and giving them their "assignments" while also dispensing wisdom to the week's protagonists. Her character is also frequently a source of the show's humorous lines, softening her portrayal as an authoritative African American woman and perhaps therefore making her more palatable to a middle-class audience. Monica (Roma Downey) is the "angel of Truth," usually distilling the core questions the protagonist must answer and calling upon him or her to make the "right" decision. She functions dramatically as the innocent woman-child, a lumanescent beauty with lilting Irish accent, guileless in her interactions with others. Downey's off-screen persona as a diminutive, contented mother of an infant is used to reinforce her role as the program's desirable yet unavailable woman (See, i.e., Powell, 1997). Andrew (John Dye), the "angel of Death," assists persons and their families in the passage from this world to the next, and functions as a helper to the other angels when death is not imminent. His all-American good looks, soft-spoken earnestness, and furrowed brow of concern for those he is assisting has made him the idol of the show among teen girls, although his character, in contrast to Downey's, is rarely romantically linked with "humans."[6] The popularity of the show has catapulted its stars into an interesting form of celebrity which has capitalized on their spiritual identities. The fact that Della Reese is also a minister in "real life" was initially employed in press accounts to suggest that her nurturing role as Head Angel carries

[6]Mirroring "real life," however, one episode in the fall of 1997 depicted him as the object of a young teen's adoration.

over to her behind-the-scenes persona.[7] Moreover, Downey relates a story of recognition in which a woman and man walking down the street stopped her. "Are you that angel on TV?," the woman asked. When Downey nodded in affirmation, the woman turned to her husband and said, "*You* see her too, right?"(Arnold, 1997).

The episode employed for the per-led discussion groups was titled, "At Risk." It originally aired in late April, 1997, in conjunction with President Clinton's Summit on America's Future, which emphasized the importance of volunteering. The episode was introduced and closed by General Colin Powell, U.S. Army (ret.), who served as the General Chairman of this event. His remarks, underscored by a slow hymn, were filmed in front of Independence Hall in Philadelphia and called the audience to a renewed commitment to volunteering, noting that "the strongest resources we have to keep America strong are Americans themselves."

The story begins as Jason DeLee, an African American teen, breaks into and steals a car. When the police surround him, he crashes the car, rushes out with gun in hand, and is arrested. His mother is called to the police station, and coldly relinquishes her parental authority over him. This sends Jason into the custody of Juvenile Services, where he quickly enters into a rivalry with a Latino teen, Ray Estes. Meanwhile, the gruff Juvenile Services camp director is displeased when head angel Tess informs him that she will be overseeing an experimental work program for the boys. Angel Monica then arrives and is introduced as the coordinator for the work program. She asks for volunteers, although

[7] Her responsibilities as a minister, however, became a topic of debate with regard to her contract negotiations with CBS in the fall of 1997.

Jason does not raise his hand. Jason's mother then visits Jason, but he is angered when she quietly tells him that she is only bringing him some underwear and car magazines. Tess then says that Jason has been selected for the work program, and Jason's mother encourages him to "take whatever opportunity they give you," although he refuses to respond and departs. Tess encourages Jason's mother, telling her that she "did the right thing." Upon learning that Estes will be volunteering for the program, Jason decides to participate so that the two of them can "get it on." The angel Andrew drives the small group of boys to a home for disabled children, where Jason is paired with Kelly, a caucasian girl with cerebral palsy who cannot speak. Kelly's mother (also caucasian) voices concern about having a "gangbanger" assigned to her daughter, but the angel Andrew convinces her to give him the benefit of the doubt. Jason, however, is unresponsive to Kelly, until he learns that she also likes his car magazine. The rivalry continues between Jason and Estes, and when Jason demonstrates his interest in Kelly and the other disabled children, he is teased by the other boys who tell him that"you can't let nobody get to you like that." He then escapes from prison. Monica eventually finds him in the Porsche dealership as he is about to steal a car. As she stands, the strong backlight frames her ethereally, and soft music underscores her words. She reveals herself as an angel and encourages him to stop trying to run away from God, leaving him in tearful reflection. Back at the children's home, Kelly's mother angrily confronts Jason's mother, who has arrived with Monica to see where her son worked. Jason's mother responds by describing her deprivations and her love for her son. Kelly's mother responds with a tearful apology that she, too, understands difficult

times. Then Jason arrives, and tearfully notes that by learning to care for Kelly, he has learned to care for others in his life, as well. Kelly then welcomes him back, and the show ends with close-ups of the mothers, the angels, Kelly, and Jason in tears. At the close of the narrative, Colin Powell offers concluding remarks, commenting that this story was based on a true story.

In the tradition of semiotic analysis, it is often helpful to point out the discrepancies of the media text, at least as they appear to me as an analyst. These shed light on the intended audience, and also provide clues as to the "preferred reading" of the text: that which might be described as the producer's intended meaning encoded into the content of the text (Hall, 1980). Several codings reveal that audience members of the middle class are invited to identify with the characters. For instance, although we know from the dialogue and from the one exterior shot of Jason's home that Jason and his mother are poor and live in the inner city, Jason's mother is "coded" as middle class: she wears attractive clothing, silver earrings and makeup, and her hair is neatly coiffed. She is quiet and reserved in manner, yet articulate in her expressions. She is respectful and submissive with Tess, the head angel in a position of authority in the prison which holds Jason. In fact, as each of the angels are allied with the prison through their human positions, the program reinforces the idea that one can best succeed when working within the current system of justice. Moreover, there is no evidence of a support system for Jason or his mother aside from that given by the angels, which reinforces the position of the poor as victims subject to the compassion shown them by members of the middle class. The poor are symbolically held in their

(nonthreatening) place while Colin Powell calls upon middle class American viewers to have compassion and volunteer. The program suggests that the system can be trusted to be just, particularly when individuals contribute to it through their volunteer efforts.

There are also gender implications for this particular episode. The middle class coding of Jason's mother symbolically allows space for the reconciliation between her and Kelly's mother, who is also coded middle class yet has expressed race-based "hatred" by criticizing Jason and his mother. The exchange between the mothers toward the end of the program is as follows:

> Kelly's mother: Your son is everything wrong with this world...It comes from parents who don't care, who are irresponsible.
> Jason's mother: I am Jason's parent. Just me. I don't have a husband. I don't have a education. I don't have money. But as God is my witness, I love him. And I did my best. I prayed to God for a miracle, that my boy would change. Maybe the miracle is that he got arrested, because if he found some love here, and some meaning in his life, maybe my prayers were answered.
> Kelly's mother: I'm sorry. I understand hard times. I do. And I'm sorry to be so hateful. I'm just worried about my child.
> Jason's mother: So am I.

The exchange ends in a shared smile as the music swells to suggest reconciliation: in this conversation, the texts suggests, the two women have bonded over their common concern for their children. As this concern is naturalized to suggest a universality of interests, differences of race and class are erased. Tess and Monica, the key angels of this episode, look on approvingly as Jason then returns to the volunteer program and, with tears streaming down his face, tells his mother that he missed her. Jason's rescue from a self-destructive life of crime is therapeutic: the women in his life - his mother, the angels, Kelly - have gently taught him

to care for others and by extension, for himself. Given this emotional method of rescue, it is perhaps not surprising that the other male characters in this episode function primarily as background elements.

While I have tried to illustrate why I believe this program "works" within middle class ideology, such an analysis only begins to hint at the cultural location of the program and its reception among various audiences.

Touched by an Angel : Is it Kitsch?

I have to admit that I was not a *Touched by an Angel* viewer when I began my research into religion and the media in 1996. Yet in my qualitative interviews, so many people mentioned this program that it finally dawned on me that the show represented something significant to many people, from all walks of religious life, and that analyzing its cultural location would shed light on the broader issue of public discourses of religion as they are articulated at the site of media audiences.

The program fits into a television genre previously occupied by *The Waltons, Little House on the Prairie,* and *Highway to Heaven* , all programs that wrapped nostalgic notions of family, earnestness, and goodness in a sentimental affirmation of middle-class morality. Indeed, the program's success cannot be divorced from the fact that a dearth of programs in this genre existed in the television schedules of the mid and late 1990s. Like its predecessors, many of the narratives of *Touched by an Angel* dramatically address and resolve a salient moral conflict. Every episode follows the same format: the three angels observe a person (or group of people) encountering a difficulty, and they enter that person's life

in human form as a coworker, doctor, neighbor, or in a similar role. After dispensing support and advice, the difficulty reaches a dramatic climax: the music swells and the strong backlight illuminates the angel as he or she reveals his or her true identity as an angel sent by God. In this monologue, the angel tells the person that "God loves you" and clearly articulates the choice that person must make to resolve the situation. The show ends as a resolution is reached amidst swelling music, joyful smiles, and a not infrequent shedding of tears.

In addition to the weekly program, *Touched by an Angel* has tie-in merchandise available through both the CBS Store and Christian and retail bookstores. Of course, the program's popularity is also related to the angel "trend" of the past five years. Angels, it should be noted, are among rare artifacts that can be found represented simultaneously in evangelical Christian bookstores and in new age/metaphysical book stores.

This connection to a specific popular, vernacular artifact makes it tempting to locate the program within or at least related to the tradition of Christian kitsch.[8] In its sentimentality and commercialization, it clearly shares some common ground with the commodified artifacts of inspirational music, scripture posters, and Jesus t-shirts: items that are mass produced and frequently denigrated by aesthetic critics as banal or vulgar.

Colleen McDannell points out in *Material Christianity* that there are many approaches to kitsch within the tradition of art criticism, and these share common ground with the criticisms of mediated popular

[8] Currently, there are four books available that hold some relation to the program; additionally, the CBS Store sells t-shirts and caps with the program's logo. Also, there are innumerable artifacts of angels available through Christian and metaphysical bookstores which, while not directly related to the television program, are a significant aspect of the same public discourse.

culture. McDannell cites Stuart Ewen (1988) as a critic who defines kitsch as "cheap, mass-produced imitations of elite style," those products made by members of the middle class in an effort to identify with the upper class. While she terms his a "cultural" response, she notes that other critics employ an "ethical" response which suggests that the producer of kitsch is, as in one such critic's words, "an ethically base being, a malefactor who profoundly desires evil (Broch, 1968, p. 76, cited in McDannell, p. 166)." As McDannell characterizes this position, "Kitsch does not reflect the true, real world but instead leads people to believe that kitsch alone could satisfy fundamental needs (p. 166)." Thus the infinite is "sentimentalized," making a relationship with the divine seem more accessible than it is - or at least, more accessible than some theologians would prefer to think it is.

Peter Williams (1980) articulates a similar position toward popular culture in his *Popular Religion in America.* He observes that the sentimentalism characteristic of the popular quasi-religious literature of the 1970s involves

> a certain fuzziness of contour, a blurring of sharp distinctions in the interest of uttering 'beautiful thoughts' the precise import of which is somewhat elusive...The general effect is that of a sentimental mysticism, designed to evoke vaguely religious thoughts and emotions but having roots in no coherent tradition (p. 210).

Critiquing Erich Segal's popular novel *Love Story,* he writes: "The immortal message of the book - 'love is never having to say you're sorry' - is exactly the sort of unfocussed, ill-conceived 'beautiful thought' which is characteristic of the sentimental tradition (p. 210)." He bemoans the lack of "conceptual rigor" in this and other popular expressions.

In some ways, these "ethical" approaches to cultural studies may be

linked conceptually to the long line of cultural criticism found in Matthew Arnold's (1960) nineteenth century writings. Arnold defined culture as "the best that has been thought and said in the world (p. 6)." He feared that "culture" was threatened, and warned of the dangers of the emergent "anarchy," or popular culture, which he believed was being ushered in by the popular working class practices of "their beer, their gin, and their fun (Arnold, 1954, p. 591, cited in Storey, 1993, p. 26) ." As Turner (1990) points out, the "culture and civilization" tradition has been most clearly defined by its critique of the mass culture of the twentieth century, including the mass distribution of "penny dreadful" novels, women's magazines, cinema houses, and television. The concern over the moral and aesthetic content of these mass distributed media at mid-century was that their consumption would lead to a debasement of social life. There was an unapologetic elitism in these criticisms, as they were lodged during the historical moment of the rise of autonomy among the working class.

From Kitsch to Practice Theory

Bourdieu (1977) has argued that the denigrations of kitsch and other forms of popular culture serve a specific ideological function: by deeming some artifacts as more appropriate or progressive than others, the tastes of the more elite group in culture are legitimized. Taste, Bourdieu points out, reflects economic and educational levels. Thus any critiques of popular culture must read within a consideration of the position and status of the critic. Furthermore, Bourdieu's theoretical work offers an approach to culture and cultural artifacts which is distinctive from the "high culture/low culture" analyses outlined here. Introducing the

concept of the *habitus* , Bourdieu has argued that individuals negotiate between social structures and individual habits in a dialectical process, and that culture is best understood through a more *practice-* oriented approach. Bourdieu's approach asserts that the system - including its social, economic, political, and I would add, gender aspects - exerts a powerful, determining effect upon human agency; however, we can only understand how the system continues to function (or is subverted) by exploring individual actions within it.

I recognize that I am assuming a rather direct relationship between a television program and an artifact of material culture. This may seem an unlikely comparison; after all, one can resist purchasing a ball cap with the statement "Jesus Christ...He's the Real Thing" by avoiding the locations in which one would purchase such an item, but one cannot avoid television's ubiquitousness. We assume that the Christian bookstore is closer to the realm of the private sphere in that it serves private interests, whereas television is the contemporary *locum tenens* of the public square: essentially, we expect more from television in terms of its ability to represent a pluralistic society to itself. In a Christian bookstore, in contrast, one purchases what is potentially meaningful and leaves the other materials on the shelf.

I would like to suggest that this dichotomy between artifact and television rests upon an assumption of the media as something that is more directive and powerful, or at least more sinister, than an inanimate object offered for sale. Television is frequently viewed as a "culture industry," manufacturing manipulative fare which diverts attention away from social problems and affirms the status quo (See, e.g., Adorno, 1991).

Yet with the proliferation of channels and other media offerings, it is clear that the media industries are far from monolithic. Further, viewing the television industry solely from the perspective of its economic functions does not adequately explain its cultural role. As Fiske (1987) notes, symbolic goods such as television programs do not have a clearly defined *use-value,* but they do have an important *exchange value* in the cultural realm. As he points out, "What is exchanged and circulated here is not wealth but meanings, pleasures, and social identities (p. 312)." He points out that cultural commodities such as television programs are material commodities when they are sold by producers to distributors, but then become producers themselves as they generate audiences which are, in turn, sold to advertisers. Audiences in turn, he argues, shift from being a commodity to a producer of "meanings and pleasures." Once the audience members cease to produce meanings and pleasures, they withdraw from the equation, leaving the distributors and producers with the task of locating a new cultural product with which they can generate a new audience. The statistics demonstrating the high failure rates for new television programs as well as for new films, musical recordings, and other media point toward the ability of media audiences to resist that which holds no meaning for them.[9] Much like artifacts, therefore, it is difficult to control or predict which cultural products will be successful in provoking meaning among its potential recipients. For this reason, studies of the popular in media as well as in art, following Bourdieu, have turned to the examination of the *practices of interpretation* , exploring how and in which contexts cultural products come to have meaning.

[9] Fiske (1989, p. 31) notes that between 80 and 90 percent of new films fail, despite extensive advertising. Likewise, Frith (1983, p. 143) has argued that 80 percent of single releases and albums lose money.

This, then, has implications for the ways in which we might understand the relationship between popular cultural forms and religion, or the emergent "religion of the media age," as Hoover (forthcoming) terms it. Hoover (1997) has argued that symbolic practices are now seen to be in the hands of *individuals* , who construct "sets of cultural symbols into unique and personally salient patterns (p. 285)." Following postmodern critic Frederic Jameson, he argues that this appropriative process moves in two dimensions. It moves across *time* , as particular narratives of history are employed in the service of the purpose of the moment, and it moves across *space* , as elements from various traditions are extracted, separating them from their cultural context. Thus religious identity is constructed somewhat idiosyncratically, as various elements are brought into coherence on the individual level. I argue that while this does not imply a total disappearance of institutional referents nor of certain larger patterns of social structure and discourse within which individuals conceive of their individual experience, it does require a removal of institutions from a privileged position in the consideration of religious identity. Religious institutions are therefore *at best* equal with other symbol-producing and maintaining sources with regard to religious identity-construction in the contemporary mediated age. The media, through their cultural role as constructors of dramatic stories and news, play a key role in legitimizing and shaping certain religious discursive formations of the symbolic inventory. Thus we are returned to our current exploration of *Touched by an Angel* , a program that is clearly meaningful for a significant population in the U.S. today and thus might be expected to articulate the discourses of religion most popular among the

American public.

The Discourses of Touched by an Angel

In chapter five, I discussed 16-year-old Elizabeth's case study. At that point I termed the discourse I believe she embraces with regard to her religious identity an *institutional* discourse, as her understanding of her own religiosity is closely related to her affiliation with her local Lutheran (ELCA) church. In this chapter, I will delve into more examples of differing discursive strategies I believe were revealed in both conversations with the "case study" teens and among the discussion groups they led.

An Institutional Discourse

As noted in chapter five, Elizabeth equates her interest in church with her interest in religion. Yet despite her tendency to think of religion in relation to her religious institution, she notes that people in her church, including her pastor and youth director, rarely speak about media. Thus the framework for the way in which she approaches media is not directly informed by her institutional experience in a way that she is able to articulate easily.

While I will refer to Elizabeth's approach to religion as an articulation of an*institutional discourse,* I want to note that this does not imply that Elizabeth agrees with the theological stances of her local church, if she even knows them. Like many young people her age, in fact, Elizabeth's knowledge of and commitment to what one might wish to argue are specifically Lutheran beliefs is rather vague. While she relies

upon reference to the religious institution to draw a distinction between "religion" and "beliefs" (discussed in relation to the contrast between television programs employing overtly religious symbolism - church building, clerical collar, etc. - and those dealing in the realm of the supernatural), she is less clear on the boundaries that might distinguish "religion" and "beliefs" in her own belief system. When explaining her commitment to her local church, for instance, she relies upon therapeutic rather than theological language, pointing to the support that members of her church gave her during her parents' divorce. Similarly, her preference for *Touched by an Angel* over *Soul Man* based on its openness to "beliefs" beyond the institution also suggests the lack of distinction between her institutional affiliation and a popular religiosity. The main difference seems to be simply whether or not one has an institional affiliation. This results in a rather truncated discursive position.

This lack of a language with which to discuss religion beyond the institution also emerged in her peer-led discussion group. In their discussion of the program *Touched by an Angel,* she and the other members of her group who shared an instititutional affiliation with religion wanted to distinguish themselves discursively from their friends who did not have any institutional affiliation, yet they were hard-pressed to express that distinction. When asked whether or not they liked the show, for example, there was not a clear division which placed those with an institutional affiliation in the same camp and opposite those with no affiliation. This created a tension that emerged at several points in the conversation.

Elizabeth and her friend Lisa, who are both active in their mainline

Protestant churches, regularly watch the show, as does Tara, who has no religious background. Allyson, who also has no religious background, did like elements of the show although she was not a regular viewer. Two of the girls who only very occasionally attend church, Betsi and Vickie, did not like the show.[10] Vickie, watching the program for the first time, commented incredulously on the semiotic code during the angel Monica's revelation to Jason: "Oh, so they kind of appear, out of thin air, with the light behind them." She and Betsi also broke the mood of the sentimental ending:

> Vickie: (sarcastic) Oh, this is soo touching!
> Betsi: That's what I don't like about these shows. For some reason, everyone's always crying.
> Allyson (no religious background, agreeing): Oh yeah.
> Betsi: Always!
> Tara (no religious background, but regular viewer of the show): It always makes me cry.
> Allyson: Sorry to disappoint you, but I'm not crying.

These participants in the teen girl discussion group agreed that the main character of the particular program and his mother were like people they had known, although they did not elaborate on this. They disagreed, however, about whether or not the portrayal of the angels was "realistic." This led to an interesting exchange between Tara, a teen with no religious background but who regularly watches the show, Elizabeth and Lisa, the active church members who are also viewers, Allyson,the non-religious non-viewer, and Betsi and Vickie, the teens with a more marginal connection to organized religion who did not like the show:

> Elizabeth: Were these people similar to things or people you've known?
> Allyson (no religion): The kid was, but the angels, not at all.

[10] Vicki's family occasionally attends a Catholic church, while Betsi's family attends "maybe on Easter," and has gone to "various churches."

Vickie (marginally religious): I've never known anybody that niiiice.
Betsi (no religion): The bad people or the angel people, it doesn't matter.
Lisa (religious, viewer): Okay, yes, they all seem like people I've known.
Allyson (no religion): They all seem realistic, except for the angels.
Vickie (marginally religious): Not the people I've known, people I've heard about.
Tara (no religion, viewer): How do you know that's not real?
Allyson (no religion): Yeah, yeah, yeah, mon.
Elizabeth(religious, viewer): It's kind of a belief thing.
Tara (no religion, viewer): What (is)?
Elizabeth(religious, viewer): Like, there ***could*** be angels here.
Tara (no religion, viewer): (agreeing enthusiastically) Oh, of course!
Allyson (no religion): (sarcastic) Like angels are gonna show up, and all of a sudden, poof. I don't believe that.
Elizabeth (religious, viewer): It could happen.
Tara (no religion, viewer): Okay, I might be an angel, people! (laughter)
Lisa (religious, viewer): Hello? ***I*** am the angel, okay?
(more laughter)
Tara (no religion, viewer): ***You're*** an angel?!
(laughter)
Lisa (religious, viewer): (more serious tone) An angel came to Mary! An angel came to Mary.
Tara (no religion, viewer): (agreeing) Yeah, an angel came to Mary.
Vickie (marginally religious): How would an angel find us, exactly?
Allyson (no religion): Well, angels usually know how to get to other peoples' houses (making a reference to Lisa's mistake in finding Elizabeth's house).
Lisa (religious, viewer): So I was a street off. Sue me.
Allyson (no religion): 'Maybe if we wander around a little bit, maybe we'll find it.' She's like, so we'll drive around until we find it, that's what she said to do.
(continued discussion about getting lost)

Here we can see the teens with some religious background struggling to find some way to articulate their belief in angels, although Tara, the non-religious viewer, seems to share the same general knowledge as the girls active in their churches. Lisa hits upon an illustration of an angel from the New Testament Christmas story, but this reference is perhaps unclear to those with no religious background. More likely, however, it does not

add to the argument for the existence of angels. The girls then turn to another topic of conversation. In the course of this discussion group, the girls never mention their different positions with reference to institutional religion, although, as we will discuss at a later point, they all do agree on several of the more general points of popular religiosity. We will return to a discussion of the ways in which the program echoes this institutional discourse and its interpretation by others who identify with this approach to religion. First, however, let us consider three other approaches.

A Revelatory Discourse

Michael, as introduced in chapter three, is an African American fifteen-year-old in a single parent family of modest means. He is a good student and an active participant in a neighborhood community center, where he has honed design skills for the World Wide Web. In contrast to Elizabeth's truncated discourse on religion in her family context, Michael's family and friends frequently speak of religious issues and beliefs. Michael has followed his mother's trek from Baptist to African Methodist Episcopal to Church of God in Christ, although he is less involved in his specific congregation than Elizabeth is in hers. His mother, who is very articulate, weaves the language of Christian salvation and conversion into her life story and often uses religious language in their everyday home life. Like his mother, Michael uses a more expressive and personal language to discuss issues of faith:

Lynn: Do you think of yourself as religious?
Michael: Yeah.
Lynn: So, what does that mean to you, to be a religious person?
Michael: It means that - (chuckles) I don't really know. I just believe in

God, that he's my savior.

Of the five case study teens, Michael is the most comfortable speaking with his friends about issues of faith, although he cautions that he does not do this frequently. He (and his friends) relies upon what I will term a *revelatory discourse* : one that emphasizes God's love as it is revealed in the world through persons and supernatural beings. The individual's response to such a revelation is emotional, and is meant to result in the embrace of a moral life. As we will see in his group's discussion, this revelation may or may not be prompted by or occur within a religious institution.

Like the conversation with Elizabeth noted in chapter five, some of the most illuminating parts of our conversation emerged in Michael's comparisons between television programs that have some religious referents. For example, he offered a differentiation between *Touched by an Angel* and *7th Heaven,* a family program on the WB network which might be considered in the same post-*Waltons* genre as *Touched by an Angel. 7th Heaven* features a white middle class minister and his family of five, three of whom are teens. Although it is popular with teens, Michael describes it as "fake," particularly when contrasted with *Touched by an Angel:*

> *7th Heaven* is like a soap opera to me. They have soap opera problems. But I think...*Touched by an Angel* tells you really about God and how he feels about people, but *7th Heaven,* it just doesn't touch bases with that as much as*Touched by an Angel.*

Michael's comparison of *Touched by an Angel* to *7th Heaven* provided me with an opportunity to explore how his views of the portrayal of institutional religion might be similar to or different from Elizabeth's

approach. I began by asking him why he felt that *7th Heaven* seemed "fake," to which he replied:

> Michael: All the problems in *(7th Heaven)* are just, out there. And then the ways they solve them are just like, 'oh, the family loves you, okay, I'm okay now.' It just, doesn't get me.
> Lynn: Do you know some ministers, how they spend their time, in terms of how much time they spend in church and stuff?
> Michael: Yeah. They spend a lot of time in church.
> Lynn: Do you think *Touched by an Angel* would be better if they spent time in church? 'Cause they didn't spend any time in church on the show that you just watched.
> Michael: Yeah, but - . See, what I'm saying, like, at least they say something about *God*. See, on *7th Heaven*, they don't say about God, they say about *the minister*. Like, 'it would be better if you talked to me 'cause I'm a minister.' But at least when they talk to an angel, they talk about God.

Like Elizabeth's contrast between *Soul Man* and *Touched by an Angel*, Michael discusses the differences between *7th Heaven* and *Touched by an Angel* in terms of the markings of institutional religion in the form of the minister: if there is a minister, it is somewhat "set," as Elizabeth said, as a religious program. Yet Michael's distinction also suggests that he finds the institutional discourse of religion lacking; *7th Heaven* is not appealing because although it gives reference to a church, it does not reveal anything about God; *7th Heaven* does not "tell you really about God and how he feels about people." This suggests that the institution alone does not wholly reveal God's presence in the world. *7th Heaven* also does not speak to the "real-life" problems of teens in the inner city, as *Touched by an Angel* does, at least occasionally.

In fact, in contrast to the girls' group, the boys in Michael's group (all of whom were African American) had no trouble conceiving of the characters, the situation, or even the angels in the episode of *Touched by an Angel* as "realistic." Upon first seeing the episode I'd chosen for the

group to review, for example, Michael told me that he had seen the episode and asked if I knew that it was supposed to have taken place in Denver, the city in which this study has taken place. He then pointed out the Denver police uniforms and logo, which had been completely missed by both other focus groups, another teen viewer Jake, and by me. When asked whether the people in the program were similar to others they knew, one member of Michael's group, Dillon, replied: "Well, my older brother, he's kinda like Jason, but he doesn't help anybody, so - ." Dillon's older brother, Michael later informed me, was currently in prison and had, like the lead character in the episode, served a term in a juvenile detention center a few years earlier. Michael also noted that he and his friends have known several boys from school who have stolen cars, although he thought it was implausible that Jason would have been "stealing cars since he was twelve" as the story claims, as that was "too young." Perhaps most striking, however, was in the comparison they drew between Head Angel Tess and their own mothers:

> Michael: Do you know people who act like someone in the film?
> Randall: Yeah. My mom. She's an encourager. She'll tell you to keep going at it, don't give it up. Really encouraging. [...]
> Daryl: I'll go with Randall's comment. My mom is the same as his, she encourages us to do the right thing, and she encourages us not to lie and to tell the truth.

They also all agreed that their mothers would approve of the program for its moral encouragement:

> Daryl: My mom would just love everything about [the program], 'cause it shows, it teaches us young people that, the way to go is the*right* way.
> Randall: My mom would just love it. She would have no grief, nothing - she would just like it, straight up like it.
> Michael: What were you going to say, Burt?
> Burt: I was about to say, I don't think my grandmother, what she wouldn't like about the [program] is, that lady that kept harrassing

Jason, she wouldn't give him a chance.

In the last statement, Burt revealed that he recognized the issue of racism as it was encoded into the program in the form of a white woman who makes harsh judgements against the main teen character but then recants. While he agreed with the program's normative stance against racism, the group did not speak of racism other than in this instance. Instead, their praise of the program was in relation to their belief that *Touched by an Angel* provides a moral compass sorely needed in their real-life experiences:

> Michael: With *Touched by an Angel* , they [the characters] are talking with the angels who come straight from God, and they're telling them exactly what God wants from them, and what God wants to say to them. [...]
> Randall: [The program is] dealing with problems that children have. They're getting them off the streets, they're teaching them something smart, putting them into jail, getting them [straightened out].

This moral message of getting young people "straightened out" through the justice system is a part of God's "message," according to these boys. In many ways, the program*Touched by an Angel* itself provides a revelatory message, as it features angels who "come straight from God," telling "exactly what God wants from them, and what God wants to say to them," as Michael says. Their mothers are also responsible for revealing the message of God's love and God's moral compass to them, as these women affirm, as the program does, that "the way to go is the right way." Perhaps this explains why both Randall and Dillon's interpretations of the program placed the main character's mother in a more central position than it seems was actually afforded by the storyline. When asked if the teenaged boy's problem could have been solved differently, for example,

they reveal their understanding of the angels as somewhat incidental:

> Randall: Well, if I was [Jason's] mom, I'd just say [to the prison officials], 'could you guys like keep him for a couple days, teach him a lesson, talk to him, stick him in a room, teach him a lesson, keep him for a couple weeks, and then it would teach him a lesson in life, like, 'don't steal.' Like, teach him a lesson that he'd need to really learn. And then he'd probably learn a lot. He'd probably never mess up again, and he'd keep his head straight into the books.
> Dillon: I really think they shouldn't change anything [in the show], 'cause I think his mom did the right thing, like, making him go to that camp.

It is interesting to note that the boys, even more adamantly than the girls (or the other group of boys), affirmed the normative approach to the justice system portrayed in this episode of *Touched by an Angel.*. Indeed, I was surprised to learn that all of the boys liked the program. As noted earlier, Michael named *Touched by an Angel* as his favorite television program. Daryl and Dillon, twin brothers and Jehovah's Witnesses, were regular viewers of the program, as was Randall, who has no religious background. Adam and Burt, whose mothers both attend church although they do not, had not seen the program before, and were less enthusiastic than the others about the program.

In comparison to the caucasian teen girls' discussion group, the African American boys were more specific in their descriptions of what they believed the program communicated about God. They infused their description of the program with a message of equality that was not mentioned at all in the girls' conversations:

> Randall (no religious background): The [program] had a message that God loves everybody. It's not what color you are...handicapped or not. God really - I'm not a racial person. Everybody's okay with me. And what really, people should be more like the *angels*, they were *thinking* . Tess -
> Daryl (Jehovah's Witness): More like *God.*

Randall: Yeah, more like that. If you come out with a loving heart, you get love back and all that, and then you'll be treated equal.
(the boys all clap)
Michael (Church of God in Christ): (sarcastically) That was beautiful. Dillon?...What did you think the show was trying to say God was like?
Dillon (Jehovah's Witness): God was kind and loving and he thought everyone - he *thinks* - everyone is special to him. He loves everyone with all his heart, and he will treat everyone the same way, he doesn't discriminate or anything like that.

Again we see the lack of distinction between those who are members of a religious institution and those who are not, although in this group no tension emerged regarding this issue. Institutional identity as a marker of distinction seemed less important among these boys than among the girls. Instead, the boys consensually praised the moral message of the program, stating that they liked that it "taught a lesson." They affirmed the norm of a just system in which the responsibility for doing right falls squarely in the realm of individual choices:

Randall: I liked the angel [Tess] that kept encouraging him to go to the [volunteering] program, and kept him in it, 'cause she was really right, he needed to keep going at it, 'cause [if] he stayed there [in the juvenile hall], he'd probably have to do more time, and it [would have] been harder on him, but as soon as he went back, he helped somebody, he really helped somebody get through. I also liked the Irish [angel] - the one that come in the car, 'cause she was also being good to him in a way. She was telling him, like, 'why would you go, like steal - God loves you. Everybody else can understand you and relate to you.'
Daryl: I liked them *all* [the angels], because they *all* encouraged him to do, do what was right instead of doing what was wrong, that got them in juvenile hall in the first place.

That, in fact, appears to be a central feature of the revelatory discourse: that a God of love, who may or may not have much to do with institutional religion, encourages young people to live a moral life so that, as Randall says, "if you come out with a loving heart, you get love back and all that,

and then you'll be treated equal." As illustrated in this discourse, God can be revealed in numerous ways, including through one's mother, possibly one's church, and also, even through a television program about angels.

A Sentimental Discourse

As noted in chapter three, unlike Elizabeth and Michael, Jake has had no experience with religious institutions at all. Like Michael's mother, Jake's mother also describes herself as religiously conservative, but she dropped away from her church when Jake was very young. While she and Jake's stepfather, a lapsed Catholic, are interested in returning to religion "someday," Jake has learned most of what he knows about religion from movies he has seen in the cinema and in English class, where he viewed a film about "the guy who parts the sea" and another one on "the guy with the coat of many colors." His class was reading the story of Job during our interviews, and he is intrigued by Job's faith in the face of incredible suffering. His family often speaks of "guardian angels" that protect the family members from harm, and this concept of benevolent supernatural beings is reflected in Jake's talk:

> Lynn: Do you think of yourself as a religious person?
> Jake: I guess in some ways, probably, 'cause I do believe in God and all that. I don't really believe there was an Adam and Eve, I guess. I don't know how we came to be here. I don't know, I know there has to be angels or something out there, 'cause there's been a couple times where it's been close, where you think you're gonna die or something, but then somehow you get out of it.

As his experiences with "guardian angels" inform his diffuse sense of religious identity, Jake employs a *sentimental discourse* in his discussions of religion.[11] Like Michael, there is an element of morality in Jake's

[11] I rely here on Williams' (1980) description of sentimentality in religion, as described earlier.

worldview, although this is less equated with God than with practical experiences of everyday life. He notes, for example, that his parents do not want him to watch violence on television and in the movies because of its potential ill effects, although he himself is less convinced of the reasoning behind the restrictions placed on him when he was younger. When compared with Elizabeth and Michael, Jake's interest in the supernatural has less do to with any other aspects of his life. This sentimental discourse of religion therefore allows for a warm, accepting approach to religion; he is not hostile to religion even in its institutional forms, but nor does he see much practical relevance for practising it. The role of religion is to simply provide positive, or sentimental, feelings, and this rarely interferes with more practical issues of daily life.

Jake is not a regular viewer of *Touched by an Angel;* indeed, he had never seen the program before we viewed it together. I knew that he did not have much knowledge of Christianity, yet was still surprised by the limits this placed on his ability to interpret the religious imagery in the program given the extent to which Christian imagery and language is infused in U.S. public discourse. After we had finished watching the program, for instance, we had this conversation:

> Lynn: So, what did you think of the angels?
> Jake: Uh - what do you mean?
> Lynn: Okay, that might be a hard question. Did you know, first of all, that they *were* angels?
> Jake: Uh, I'm not sure.
> Lynn: You heard [the angel] Monica's speech, right, where she had the light come down on her head, when Jason [the main character] was in the car. 'Cause you've never seen this show before.
> Jake: No.
> Lynn: (pause) Did you get that they were angels?
> Jake: Not really. Like, if they wouldn't've told that they were [angels] in the beginning, like the little introduction where it says *Touched by an*

Angel, you probably wouldn't've known.

From this exchange, I learned once again that as an interviewer, I cannot assume that my research participants will see even what seems to me an obvious point of the text. Yet interestingly, Jake did not see himself as an "outsider" to the program's Christian language. Instead, Jake discussed the angels in relation to two other broadly popular ideas of the supernatural: guardian angels, and ghosts. Angels are "always coming down to help you in some way," at least as they are depicted on television and in the movies, he explains. He sees this as consistent with his own views of both God and angels, which are similar to what he said before about guardian angels:

> I really don't think much about religion. I know there's people that believe in different Gods, and there probably is one. There's something, 'cause there's all these situations where, you don't know how you get out of 'em - like, you're about to get hit by a car or something, and then something happens, and it's like you're okay.

As an example of a mediated illustration consistent with his view, he cites the Patrick Swayze-Demi Moore movie *Ghost* - a movie with no references to angels, Christian or otherwise - but whose central character returns from the dead to help solve a crime and assist his loved one in averting danger while therapeutically working through both his grief and that of his spouse.

Thus Jake, an unlikely audience member for *Touched by an Angel,* affirmed he liked the program, and he believed that his friends would like it, as well. Jake's interpretation of the episode, however, had surprisingly little to do with the angels or their supernatural qualities. Instead, he related to the volunteering experience of the teenaged main character, and affirmed the transformative power of volunteering:

Jake: I liked how it (the program) related to real life. There's a lot of people who go out there and they just don't think anything's going their way, like that guy. And also, I can't remember what the school's like (where the main character volunteered), but it's like something like where I went (as a volunteer to work with disabled children).

Jake notes that volunteering experiences can be transforming, although he is less sure of what end will be achieved in that transformation other than a vague sense of self-satisfaction.

Jake affirmed the *sentimental* discourse he saw in the program when he discussed what he believed it communicated about God:

Jake: It said that he was watching over everyone, he **was** everyone. Even though they made a lot of mistakes and they don't think they're gonna get anywhere in life, they can.
Lynn: What do you think the show ws trying to say God was like?
Jake: Like, uh, he has no - everyone is equal to him, he loves everyone, I guess.
Lynn: Was it similar to the way that you think of God?
Jake: Yeah, probably.

Thus again, we see that the program is interpreted positively, even though the position of Jake as the viewing subject is quite different from that of Michael or Elizabeth.

Interestingly, like the boys in Michael's group, Jake saw the role of the main character's mother as central to the boy's transformation:

Lynn: Do you think that it was necessary for the angels to help him in solving the problem, or was it the fact that he volunteered?
Jake: Well, wasn't it his *m o m* that volunteered him to do the program? So I don 't know, I guess the whole idea is [that it was] the angel's idea to set up that program. So it wouldn't've even happened without them.

The angels, he conceeded, were probably necessary to the particular story as they set up the volunteering program. However, his tone suggested that he was correcting me in a misperception; it was the boy's the *mother,*

rather than the angels or the central character's actions, that were key to the boy's transformation in Jake's view. In this analysis, the angels serve a practical role that could have just as easily been fulfilled by humans. However, the fact is that they were there was "read" as referencing something vaguely positive about the supernatural existence of good in the universe.

A Discourse of Difference

As I noted in chapter three, I selected my third peer group discussion leader because, as a conservative Muslim, I believed Hasan would see himself as an 'Other' in relation to the *Touched by an Angel* program. Indeed, when we watched the program together the first time, he did seek to clarify how he viewed the angels depicted as specifically Christian and quite different from his own Muslim understanding of angels:

> Hasan: There are angels in Muslim tradition.
> Lynn: So, are the angels in the show similar to the angels you believe in?
> H: Well, angels, a lot of shows deal with angels, but angels don't come down and appear - that doesn't happen in Islam. It happened to the prophet Muhammed when one angel came down and appeared as a man, but wasn't a man. The common belief, that I see from TV, is Christianity - when you die, if you're good, you go and become an angel. But in Islam, those are two separate things. You can't become an angel. God made the angels specifically to serve him, and God made humans to - he wanted them to serve him, but he gave them the choice. Angels *have* to do it. Angels have to do that, whatever he says.
> L: So, are these Christian angels, then?
> H: Yeah.

Like Jake, Hasan saw the angels as somewhat unnecessary to the story:

> Lynn: Do you think that the problem could've been solved differently,

without the angels?
Hasan: Well, yeah, if somebody was nice like that [like the angels were]. They didn't really use their angel abilities. What does that have to do with angels, and - what if a *person* -'cause he believed they were persons, and everybody else believed they were persons, so a person could just do that [lead the volunteer program, help Jason to see that he needed to change]. The only time that they [appeared to be] angels was when she disappeared in front of the Porsche.

Unlike Jake, Michael, Elizabeth, and their groups, however, Hasan found the program too "sentimental" for his tastes, and also largely "unrealistic." He prefers programs about Muslims, and watches documentaries on his own faith, as well as some of what he calls the "historical" movies of the Judeo-Christian faith, such as *The Ten Commandments* and *Jesus of Nazareth*. When he and his family watch these movies, he explains, "that's the Christian belief, but lots of it are the same [as the Muslim faith], but when it differs, our parents tell us. So, it's a historical thing we believe and watch."

Much to both my surprise and Hasan's, several members of the discussion group he led liked the program very much. In fact, they all contributed enthusiastically when asked what the show communicated about God:

Hasan: What did you think this show was trying to say that God was like?
Mitch: Good, and loving -
Bailey: That God is loving -
Chuck: - to everyone.
Kenny: - and forgiving.
Mitch: Forgiving.
Mike: - forgiving, yeah.
Chuck: Forgiving's good, too, forgiving.
Was this similar to the way you think of God?
Chuck: Yeah.
Bailey: Yeah. [...]
Hasan: Was it any different than what you believed?

Bailey: No.
Chuck: No.
Kenny: No.

Still like the girls' group, at least one member, Kenny, attempted to subvert the sentimental tone of the program by highlighting certain aspects of the story line:

Hasan: What did the program say about God?
Bailey: It said that God is always present -
Chuck: - God is -
Mitch: - and he cares for us.
Bailey: - And the fact that -
Chuck: He exists.
Bailey: - he's influencing your life.
Kenny: Yeah, that everybody can have miracles. Even if they steal cars. Or are very dubious. Or smoke. And choke little girls that are cripple. (laughter)

In addition to suggesting the sentimentality of the program, it is possible to interpret Kenny's comments as also signaling that religion is, for whatever reason, not a topic that need always be discussed seriously among his peers. It also may express a general discomfort with talk of religion, which is not usually a topic of conversation among this (or perhaps most) social groups of teens.

As noted in chapter three, while each of the groups were fairly homogeneous in that they all attended the same school and participated in social activities together, this group was diverse in terms of the racial/ethnic backgrounds of its participants. Hasan, as noted, is a biracial Arab-African/caucasian, and three other members of his group were biracial: two were African American/caucasian, and one was Korean/Philippino. The other two members of the group were African American. While Hasan is Muslim, two members of his group are

Roman Catholic, one belongs to the conservative Protestant Nazarene denomination, one attends a nondenominational church, and one is not involved in religion at all.

There were three aspects of their conversations that I believe were related both to their own positions as members of varying racial/ethnic and religious groups and their consciousness of the group's overall diversity. First, like Michael's group, members of this group noted their dislike for the television program's character of Kelly's mom, the caucasian woman who accused Jason's mother of being an irresponsible parent. Two of the group members communicated their dislike so strongly, in fact, that Hasan felt it necessary to censor their talk by turning off the tape recorder. When it was turned back on, one of the biracial African American/caucasian group members commented, "I didn't like her either. I would like to knock her out." As was the case in Michael's group, the racist statements made by this character made a more lasting impression on this non-caucasian group than the reconciliation between the two mothers of differing races that followed the statements.

Second, while tensions had emerged in Elizabeth's group between those with church affiliations and those without them, this group seemed much more confident of the firm boundaries separating their various religious experiences. This resulted in a greater degree of expressed curiosity regarding the religious experiences of the other group members, as these conversations illustrate:

Hasan: Like, do you think God is merciful?
Bailey (nondenon'l): Yes, I do.
Mitch (R Catholic): God is merciful and forgiving, and loving.
Mike (no religious background): What is mercy?
Bailey: Like, you're given mercy.

Chuck (Nazarene): Mercy is like, forgiving, type-thing.

***

Hasan: Can you think of a time when you talked about religion on television in your church?
Chuck (Nazarene): Yeah.
Bailey (nondenom): No.
Kenny (R Catholic): No.
Chuck: Like, just if you watch a movie for youth group or something
Kenny: Oh, that's true.
Mitch (R Catholic): Like at catechism.
Hasan (Muslim): Catechism, what's that?
Mitch: It's like a little group -
Chuck: Youth group?
Mitch, Yeah, youth group kinda thing, where people of the same group go to a group and they have a teacher, and they review stuff.
Mike (non religious): Like Sunday school.
Mitch: Yeah, Sunday school, exactly. Only it's not necessarily on Sunday.

Like Elizabeth's group, it is worth noting that those who are asked to provide explanations of their own beliefs and practices usually have limited facility in articulating them. Yet I think it is because of the ways in which these group members uphold a tacit tolerance for difference that their conversation about the angels, rather than resulting in disagreement, quickly dissolves into consensus:

Hasan: What did you think?
Mike (non religious): I mean, it's realistic, but like -
Bailey (nondenom): If God sends his angels, they're not gonna tell you they're angels.
Mike (nonreligious): But what if some of them *act?* [i.e., act out the role of angel, rather than explicitly telling you].
Mitch (R Catholic): Yeah, exactly.
Kenny (R Catholic): Yeah.
Mitch: Good point.
Kenny: Yes.
(momentary silence)

Interestingly, in Hasan's group as in Elizabeth's, it is the nonreligious

group member who suggests the description of angels that seems most plausible to the others. Again, to some extent this reinforces the notion that there is little conscious separation of those affiliated with organized religion and those with no such background when it comes to the "popular" expression of religion in the form of angels.

The third and most noteworthy aspect of this group's conversation, however, was the degree of reflexivity this group exhibited with regard to racial/ethnic differences. Of course, it is not surprising that they were quite conscious of both the stereotypes accompanying their own racial/ethnic background as well as those of others. Yet what was interesting is that this became the fodder for a running joke that occurred throughout the group discussion which involved various persons attempting to disguise their voices and, employing a stereotype, "speak" as a person of another racial/ethnic background. For instance:

> Hasan: Did you like the show?
> Mitch (Korean/Philippino, trying to sound like Bailey, an African American/caucasian): I woulda liked it better if it was on BET.
>
> ***
>
> Kenny (African American, trying to sound like Mitch): I'd like to see Bruce Lee as an angel.

They also self-consciously "spoke" the stereotype of their own racial/ethnic group:

> Hasan: Why'd you like the black lady?
> Bailey (joking): 'Cause I'm black.
>
> ***
>
> Bailey (on the fact that his answer agreed with Kenny's): We have unity. We're black, yes.

***

Kenny: We want some soul food.
(later, trying to sound like Mitch): Yeah, why do we have to get pizza? Can we order takeout Chinese?

There was a single occasion, however, when a comment made by Hasan was interpreted as informed not by humor but by prejudice. In this case, he was harshly sanctioned:

Hasan: Do you know anyone like Jason, or are you like Jason?
Chuck: Which one was Jason?
Mike: I know someone like Jason.
Hasan: You know someone like Jason? 'Cause you're from [west L.A.] California, Mike, right?
Mike: (loudly, challenging): What's *that* supposed to mean?
Bailey: That there are a lot of car jackings out at [Mike's former neighborhood].
Mike: (still challenging): What do you want me to say about that?
Bailey (assertively): Yes, that's all.
(pause)
Hasan (to the group): So, do you know people like Jason?
Bailey (harshly): He answered the question yes, and that's all.
Hasan: All right.
Mike (angrily): You won't just leave the hell alone about it.
(pause, change of subject; tension dissipates as they answer the next question)

When we were reviewing the transcript together and came to this point, Hasan noted apologetically on his comment: "That was kinda prejudiced." He also noted that while he and his friends often joked about their racial/ethnic backgrounds ("sometimes they call me Saddam Hussein and stuff," Hasan says), they do not appreciate it when those outside their group make the same comments, and will "stand up for themselves." Based on the fact that this incident occurred fairly early in the group discussion and the humor and ease returned soon after, it seems that when friends make the mistakes of prejudice, they can be forgiven more

easily than those outside the group might be.

I found the self-consciousness of their racial/ethnic construction particularly interesting for this project, and will comment in more detail about this in the concluding chapter of this dissertation. At this point, however, I note that when compared to the other groups and individuals who viewed the program and then discussed it, there is not a single discursive strategy that seems to be embraced by all of the members of this group. There are comments that suggest that members might view this and other programs with religious content quite differently, such as Chuck's comment that *Nothing Sacred,* the program about liberal Catholicism is "sacreligious," as is *The Simpsons,* which other group members seemed to enjoy without offense. Also, while this group agreed that they liked the program, none of them had seen it before and based on their comments I would be surprised if any of them would watch it again. Thus what I believe is worth pointing out with regard to their embrace of discursive positions is their expressed need to communicate racial differences as they perceived them in the program and among each other. Clearly they are situated as persons who are 'Othered' by popular culture portrayals as well as in their school and other social settings. Thus this discursive strategy is noteworthy in that the reconciling or assimilative approach to race in *Touched by an Angel* is subverted in favor of what is more common among the group in other contexts: the representation of boundaries, prejudice, and finally the embrace of difference.

Other Possible Discourses of Religion

I am aware that there are a number of other possible discursive

strategies that are not represented in this early study. For example, in some ways Michael also echoed a *moralistic discourse* which I encountered in numerous other interviews for this study among both persons with a religious background and those without it. This seems to be a widely-accepted discourse of religion, at least among persons in my sample, which clearly surfaced in the parents' groups discussed in the preceding chapter. As noted earlier, Michael's narratives, and those of his group, rely heavily on this discursive strategy, although future research may demonstrate that their revelatory notion of God may be less common, particularly among those with no religious background.

Furthermore, a *mystical discourse* emerged in yet other interviews conducted, as noted in chapters two and four. This discourse stresses the inexplicable power of the supernatural over the known universe. Individuals employing this strategy, perhaps not coincidentally, saw themselves at the mercy of larger systems - both economic as well as spiritual - over which they had no control. They spoke of demons, aliens, the paranormal, and other experiences, fascinated with the possibilities of a source of power that might explain the otherwise inexplicable (and largely unjust) situations in which they found themselves. Jake's somewhat interchangeable references to God, angels, and ghosts is one characteristic of this discourse, although he did not fully embrace the position of social powerlessness as did the others employing this discursive strategy who were discussed earlier. Future research, I hope, will describe this discourse more fully.

As discussed in the previous chapter, there also may be a discourse of *religion as ultimate meaning*. As I suggested there, it is possible that

this theme is more difficult to express in narrative form. However, it is interesting to me that this never surfaced in conversations with anyone other than a pastor and his family, and was not echoed in any of the teens' groups. Again, this warrants further study.

Additionally, based upon the work of Robert Bellah and his colleagues (1985), one might expect a *social justice discourse* which would emphasize a need for social change and the role of either the individual or religious group in bringing that about. While there were two families in my study who employed this language, it was limited to the parents and did not emerge clearly in their childrens' discussions, nor did it emerge with relation to the specific media text of *Touched by an Angel* or any other mediated texts, for that matter.[12]

Finally, an *evangelical discourse* was articulated by some in the study. I have not chosen to analyze this fully at this time because many previous studies have explored this approach to religion, whereas the other strains of popular religious discourse, including the *institutional, revelatory,* and *sentimental*, which may or may not intersect with evangelicalism, have been underexplored.[13]

The "Broad Truths" of Religion[14]

While I argue that the institutional, revelatory, and sentimental religious discourses which emerged in these conversations with teens and their friends may overlap or incorporate other elements yet to be explored,

[12] I should point out that with such a small sample size, I do not wish to claim that based upon this finding, we might conclude that the discourse of social justice is less easily reproduced socially than other discourses. This is an important area for further study, however.

[13] For an excellent analysis of evangelical discourse, see Balmer, 1989.

[14] I borrow this term from Hoover, S., Venturelli, S., and Wagner, D., 1995.

they each represent a somewhat different approach to religion. The program *Touched by an Angel* is open to each of these interpretation strategies. In this section, I argue that those who embrace the differing discursive strategies noted above also share an appreciation for the program based on its ability to echo certain broad themes of religion. These include an affirmation of tolerance for other religions, a universal understanding of God, and a belief that religion is an individual choice best made and expressed in the private realm.

Among those commenting in interviews and online, perhaps the most common word used to describe the show was "open." The "openness" of the show was an important part of its appeal for those in each of the discursive positions mentioned above. Elizabeth articulated this position, as we noted earlier:

> I mean angels, you don't even have to believe in God. So you can just believe that there are angels...[Touched by an Angel] can be any religion, and it doesn't have to be a religion, it can just be a belief.

Yet while this program is "open" to the differing ways of approaching religion noted above due to its polysemic text, discussing just how "open" the program is to religious pluralism further develops our understanding of the popular religiosity expressed in religious discourse. As witnessed in the various interpretations of the program, in order for an audience member to find the program enjoyable, he or she had to find resonance between his or her beliefs and what he or she understood to be the program's message. A program like *Touched by an Angel* is interesting in its ability to appeal across various lines of the "culture wars" in contemporary religious life in the U.S. (See, i.e., Wuthnow, 1988, and Hunter, 1991). Even audience members with no religious background like

Jake need to be able to recognize in the program something of their own assumptions of the nature of religion. On the other hand, those with an affiliation with Christianity recognize that the program does "quote" Christianity, and many audience members who were affiliated with Christian traditions recognized the subtle nods to Christianity in the program's imagery and language. One caucasian woman mentioned the dove, which she recognized as a traditional symbol for the Holy Spirit, while an African American woman and Ellen Gray, a caucasian woman married to a pastor, noted the paraphrased expressions from the New Testament, such as "once he was lost, but now he's found." In this way, those lodged in *institutional* and *revelatory* discourses, identifying with the Judeo- Christian tradition, recognize themselves as the intended audience for the program (as do some of those lodged in the *evangelical* discourse).

In fact, some Christians complained that the program did not "go far enough" as it did not mention Jesus Christ or a "plan for salvation." Apparently this question was raised so often that a response was posted on the program's web page FAQ[15] sheet:

> While Martha Williamson [the show's Executive Producer] shares your personal beliefs, CBS's mandate for our show is that we entertain on a GLOBAL scale. This is why you won't find the plan of salvation mentioned. Still, we hope you enjoy the show and will continue to support us. We appreciate your comments and concerns.

This statement demonstrates that the program's promoters are aware that certain Christian audience members wish to bring the program into even greater resonance with their own views. To remain popular with this group, the program must continue to employ certain referents to

[15] Frequently Asked Questions

Christianity in its imagery and language, and also communicate to this audience that it shares their views.[16] Indeed, when the program was in its early years and was threatened with cancellation, Martha Williamson appeared on numerous Christian television and radio talk shows, including *The 700 Club,* asking for support for the program. Some evangelical and more conservative audience members who find resonance with the program and its purported "purpose" of "witnessing" praise the program's lack of reference to a specific church as an "openness" which they believed made it appealing for "nonbelievers" who will, in turn, be slyly introduced to Christianity. One college-age woman expressed this common sentiment on the program's fan club web page:

> I really like the show. I think it has done wonders in bringing Christianity to the general public. People who would never have gone to church, Bible study, etc. can watch this show and find out about an all-encompassing love. [17]

Yet many audience members praise the program's "openness" by noting the *absence* of just such specifically Christian language and imagery. For instance, no evangelical phrases such as "trust in the Lord" or "salvation from sin" are mentioned on the program, and there is no mention of Jesus Christ. Nor is there any reference to the institutional church, as the program generally avoids symbols of church buildings or its clergy. Indeed, the program's "openness" is read by some, interestingly enough, as a *contrast* to the more overt forms of Christianity which some of its audience members embrace and for which they believe the program serves as an introduction. "I think this show is more what I think God is like than like, the Bible," stated Tara, the teen with no religious

[16] See, i.e., the interview in which Executive Producer Martha Williamson discusses "committing her life to Christ" in Brill, 1997.
[17] http://members.aol.com/halofan25/dmarti1.htm

background who is a regular viewer of the show, during Elizabeth's discussion group. When her friends laughed at this supposed impropriety, she protested, "No, but seriously! I like the show's portrayal of God, and I think that's more true, that He just likes people, and he loves them, and stuff like that - " "- Rather than we have to be sin-free," stated Allyson, another non-religious teen viewer, finishing her sentence. "Yeah. 'Cause I think like all religions are good," concluded Tara, equating the lack of a doctrine of sin - a key tenet of evangelicalism - with tolerance for a plurality of religious expression. Thus, one might conclude that the program's "openess" may primarily reflect its ability to embrace a universal notion of God that is not consistent with certain orthodoxies of Christian beliefs.

Similarly, the boys in Michael's discussion group interpreted the "message" of the program as one of universal love which translates into the need for tolerance of religious pluralism, as expressed in this comment:

> Randall (no religious background): [the message is], everybody's equally loved by God. And we should turn our love to God, we should love other people, it does not matter what their religion is. Everybody's equal, 'cause really, they're all different names for God, like Jehovah. I think in English his name is Yahweh, and we call him God.

The girls in Elizabeth's group had a similar response. When asked what they thought the show was "trying to say God is like," the girls replied:

> Vickie (marginally religious): Well, the show was trying to say that God is forgiving and that he, -
> Tara (no religion, regular viewer of the show): Giving, and loving.
> Vickie: Loving, yeah.
> Betsi (marginally religious): Yeah, he's like an angel.

In its ability to speak Christianity without the trappings of institutional

religion, therefore, the program fits well with the widely acknowledged declining cultural authority of religious institutions and the public affirmation of personal autonomy over issues of faith (Roof, 1993; Hammond, 1992). It is simultaneously religious and not-religious, depending, of course, on how one defines those terms and how one positions oneself in relation to those definitions. This explains why, when Oprah Winfrey discussed the program with the cast members and executive producer Martha Williamson, she could emphatically state, "(On *Touched by an Angel)* , you're not really talking about *religion* as much as you're talking about being in touch with *spirituality.*"[18] It is not difficult to argue that the "spirituality" of which Oprah speaks has much more in common with popular forms of Christianity in its universalism, appeal to tolerance, and privatization of expression than with the "spirituality" of other world religions.

Conclusion

In this chapter, I have argued that the success of *Touched by an Angel* may be attributed at least partially to its ability to speak meaningfully to subjects that are positioned quite differently in terms of race, class, and socioeconomic status. It also has resonance with various persons from across the spectrum of religious beliefs. Those who are lodged within an *institutional* discourse enjoy "insider" status as they recognize certain symbols as Christian, even as the young people whose conversations were analyzed here demonstrated difficulty in articulating a difference between their own stance and those of others who do not share their institutional affiliation. Meanwhile, those within a *revelatory* discourse such as the members of Michael's group may find the program

[18] The Oprah Winfrey Show, June 5, 1997.

itself to be a means by which God's message is revealed. Even those in Hasan's group, where no single discursive strategy emerged in relation to religion, the group members consensually agreed that the portrayals of God in this program were similar to their own. Surprisingly, even those with a marginal relation to formal religion such as Jake might find the program meaningful in its appeal to the *sentimental* and benevolent ideas of God, angels, and the supernatural.

A key question therefore emerges: is it possible for a program like *Touched by an Angel* to be simultaneously "not religious," as its producers and cast incessantly claim in the popular media,[19] and yet Christian? I believe that the conversations about the program reviewed in this chapter reveal two characteristics concerning how religion is understood in the U.S. context currently. First, there is a felt need to denigrate the term "religion," or at least to approach it as something which is only understood in the negative, as that which it is *not* : spiritual, authentic, or personal. This narrowing of the term's referent to only that which is institutional, and often only the worst aspects of institutional religion at that, suggests the decline of the mainline Protestant and Catholic denominations across U.S. public life which has been documented elsewhere. Yet at the same time, claiming that a program like *Touched by an Angel* is *not* about religion denies its Christian emphasis and thus actually *reinforces* a Christian hegemony. The fact that this "spirituality" or "openness" exists quite apart from institutions certainly makes it less coherent as a religious belief system, and many

[19] See Powell, 1997. Also, on *60 Minutes* (1997), Executive Producer Martha Williamson asserts that she does not have a "religious agenda" and even "dislike(s) the word 'religion," while also affirming that "God is the star of the show," and that the program's central message is "God loves you."

scholars of religion have pointed out the divergences from "orthodox" Christianity that are represented in the popular acceptance of such things as angels. However, the tendency to take some form of Christianity as the "natural" and universally accepted state of religion remains an important, if underacknowledged, aspect of the public U.S. worldview. As one industry critic wryly asks, "Think you'll ever see a show that celebrates Judaism or Buddhism (or, hell - atheism) as forthrightly as this series does Christianity? *Touched by a Satanist* would have a better chance of making the fall schedule (Tucker, 1997, p. 47)."

I have focused here primarily on the comments of the teens with whom I have discussed the program. It is interesting to note that most of the teens, who come from varied backgrounds with relation to religion, favored the *Touched by an Angel* episode for much the same reasons that Oprah and her audience seemed to: while coming from a perspective clearly informed by Christianity, this program is believed to be more "open" to those with faiths outside of Christianity. Thus it allows those who hold views informed on some level by Christianity to believe that in their embrace of this program, they are embracing a pluralist view of religion, as well. It reassures both those who have withdrawn from and those who still move within Christian traditions that their beliefs are natural and acceptable to all, while simultaneously denying the possiblity that other religious belief systems might approach things quite differently. In its embrace of the more generic "broad truths" of religion, *Touched by an Angel* enables its audience members to enjoy the emotional satisfaction of the devotional experience of having gone to church, without really having to go to church.

This analysis of the public discourses of religion, therefore, in some ways extends and ultimately challenges the thesis put forth several years ago by Stephen Carter (1993) in his influential text, *The Culture of Disbelief* . In that work, Carter argues that religion is trivialized in American public discourse primarily because the political Left ceded the moral force of religious language to the political Right. Carter has argued that due to the historical fear of religion's intrusion into public life, persons of faith have been encouraged to regard "God as a hobby (p. 23)." This has resulted in a truncated ability to discuss moral issues and their roots in religious beliefs, which Carter argues hampers the prospects for a pluralistic society of democratic decision-making. As Hoover and his colleagues (1995) note, the limits of the public prospects for religion are rooted in the privatization of American religion, a process first begun with religion's disestablishment at the beginning of the nation's history. Certainly this present study illustrates the limits of religious discourse and in particular the difficulties some young Americans have in discussing either their own religion or its distinctions when compared with other religious belief systems. Yet rather than witnessing the trivialization of religion in the public realm of prime-time television, however, the 1997-1998 television season has actually demonstrated a new appreciation for religion in public, or at least its ability to resonate with "middle America" and hence garner ratings for programs.

Carter, like producers and fans of*Touched by an Angel,* presumes that the religion of which he speaks and which he advocates is a "religion-in-general," as Fallers Sullivan (1995) notes. Yet just as we must question the extent to which *Touched by an Angel* is really an expression of

"generic" religion, we also note the specificity of religious discourse implied in such writings as Carter's. Fallers Sullivan makes a similar point:

> Both Carter and Clinton seem to speak in favor of religion, religion-in-general, while they both, in fact, advocate a certain kind of religion, their kind...The public discourse about religion in this country is strained because the common understanding of religion, a product of Enlightenment skepticism and nineteenth-century Protestant piety, is out of alignment with people's experience of religion (p. 78).

Religion has been defined in public discourse within a Protestant model which stresses pious devotion, rational decision-making, and a personal confession of faith - two themes, not coincidentally, that figure largely in the appeal of *Touched by an Angel.* Yet would the introduction of this Protestant language truly reflect a multicultural and pluralistic society? Moreover, how do our conceptions of public discourses of religion along these lines prevent us from considering alternative approaches to the public role of religion?

The current study of religious discourse in the public therefore suggests that the study of such discourses should not be limited to the *governmental functions* religious discourses may play within society. The *institutional, revelatory,* and *sentimental* discourses discussed here suggest that religious discourses may serve many cultural purposes and even play a key role within personal identity-construction, all of which should inform any discussions of religion's role in public and civic life. We need more information about how these discussions play out in various forms in order to truly understand how religion already plays a role in public life and to consider its prospects for a multicultural future.

Chapter Eight

Conclusion

What role do the media play in religious identity-construction among U.S. adolescents today? To answer this question, this project has argued that we must look at the *articulations* of religious identity given by teens. I have employed Hall's (1996) definition of identity as:

> the meeting point, the point of suture, between on the one hand the discourses and practices which attempt to 'interpellate,' speak to us or hail us into place as the social subjects of particular discourses, and on the other hand, the processes which produce subjectivities, which construct us as subjects which can be 'spoken.' Identities are thus points of temporary attachment to the subject positions which discursive practices construct for us (pp. 5-6).

Addressing this "suture point" in this project, I believe, implies two separate analytical steps. First, we must look to the discourses of religion that are available in mediated texts and echoed throughout the culture, analyzing the themes and patterns that emerge and the ways in which they present a plausible understanding (or set of understandings) about reality. Second, we must explore the "processes which produce subjectivities," analyzing the various social, political, economic and other contexts that frame the individual's choices as well as the narrative patterns and practices in which such articulations might occur. Subjects, I have argued, therefore *articulate* their identity with reference - whether conscious or not - to the discourses available in the wider culture, discourses that cannot be understood apart from the historical and political economic contexts in which they emerged and for which they

continue to serve certain purposes. Because these discourses are often articulated through popular mediated texts, we can say that the media do play a role in the religious identity-construction of teens. However, the issues of determination and causality are far from settled. My research has demonstrated, for example, that discourses are not created *in* the media, nor can we say that the mediated texts fully and conclusively articulate these discourses. If the mediated text itself becomes popular, it is because it *echoes* the popularly-accepted discourses of religion (or other subjects) and in turn may then gradually contribute, along with a host of other cultural factors, to changes in these discourses. Furthermore, I have also demonstrated that we cannot predict how teens will respond to mediated depictions of religion - or for that matter, even how they will define what is "religious" or "meaningful." Thus we cannot say that the teens' understandings of and approaches to religion emerge directly from their contact with the media. However, an analysis of mediated texts does provide us with information on the discursive strategies that are available to teens, thus offering clues as to how teens might make themselves understood to others when they attempt to construct narratives of religious identity.

I began my analysis with the second of the steps I outlined above, attempting to gain an understanding of teen life by talking with a number of teens individually, developing ongoing relationships with a few of them, and exploring their conversations with both their peers and family members. I wanted to explore the factors framing the creation of subjectivities. This, I believe, required an analysis not only of their experiences in the present moment as they were related to me in

interviews and through observation, but also evoked questions concerning how these experiences differed from those of their parents and possibly from previous generations. In the next section, therefore, I would like to summarize some of the findings related to the context in which teens construct identity and draw some conclusions based upon the preceding chapters.

The Contexts for Adolescent Subjectivities

There are at least two aspects of U.S. teen life that have been discussed in previous chapters that, I believe, significantly shape the contexts for religious identity-construction among young people and the role of media in these processes: the increase in a cultural norm of reflexivity, and the flattening of various social hierarchies. The first, as noted in chapter four, is related to a wider cultural analysis offered by Giddens (1991). In that chapter, I suggested that today's adolescents are coming of age in the context of what Giddens (1991) has termed a "risk society." I noted that by "risk" Giddens means more than an increased exposure to new forms of danger:

> To accept risk as risk, an orientation which is more or less forced on us by the abstract systems of modernity, is to acknowledge that no aspects of our activities follow a predestined course, and all are open to contingent happenings...Living in the 'risk society' means living with a calculative attitude to the open possibilities of action, positive and negative, with which, as individuals and globally, we are confronted in a continuous way in our contemporary social existence (p. 28).

As part of their coming-of-age task, therefore, teens must garner the skills necessary to envision various possible outcomes to their actions. They believe that they have some autonomy in making their life choices, but are also aware of the many changes that may occur over which they have

no control: divorces, job shifting or loss, and changes in the family's financial picture, among other things, have taught them this. Their life *choices* are of course greatly conditioned by their life *chances* (the latter referring to the Weberian phrase in which work is acknowledged as an economic, and hence social, determinant of lifestyle options [Giddens, 1991, p. 82]). Yet young people today sense that they have a great deal of autonomy in this area, which I believe has resulted from the flattening of adult authority and of the authority of institutions throughout culture, and the rise of part-time employment which has given them more control over their own financial resources than previous generations had.

Living in the "risk" society, as Giddens argues, has implications for how adults relate to one another, and I believe this is true for teens, as well. Teens have come to expect, for example, that while intimate relationships with friends or romantic partners may offer fulfillment, such satisfaction may be ephemeral. This was illustrated perhaps most pointedly in Elizabeth's comments about romantic interactions on the Internet, discussed in chapter five. Elizabeth had commented that such relationships are only engaged in while they are satisfying to her; once they threaten to become "too serious" or to impede on her "everyday" life she leaves them with the sense that there is little consequence to either herself or her partner.

Relationships among teens seem to be pursued as a part of a self-reflexive process, therefore, much like what Giddens (1991) has termed a "pure" relationship:

> [Pure relationships] offer the opportunity for the development of trust based on voluntary commitments and an intensified intimacy. Where achieved and relatively secure, such trust is psychologically stabilising, because of the strong connections between basic trust and the reliability

of the caretaking figures (p. 28).

The "pure" relationship, therefore, is not necessarily constrained by the structures of social or economic life; it is justified not in reference to one's social ties, but in reference to romantic love. It is sought out and maintained solely for the gratifications it provides to the persons involved, and therefore is always organized in relation to the reflexive self who asks, "how is this relationship fulfilling to me?" Likewise, sex has been transformed from the externalized "thrill" of the 1950s to something more closely aligned with contemporary concepts of intimacy and even identity. This is illustrated in Modell's (1989) argument:

> In the 1950s love had been defined in terms of meeting role expectations. Now it was 'characterized by meeting the needs of the other through interaction, commitment, affection, and non-possessiveness.' Mutuality was no longer a theme of 'coming together' in mystic sexual union but rather of each partner enhancing the other's happiness. Each couple represented a fresh negotiation of promising but uncertain potential that would endure while each partner gratified the openly pleasure-seeking *self* of the other (p. 303, citing Kidd, 1974, pp. 28-31).

Learning of and conforming to adult "roles" in romantic relationships has come to mean less, while the realization of selfhood as it emerges in relation to the negotiation of gratification has gained in importance. Thus instead of solely looking to the media for "role models" as to how one behaves in "real life" relationships, the role of mediated texts may be better understood as echoing a model of how to negotiate self-satisfying relationships in "real life." Moreover, the media themselves should be seen as a potential site for such intimate pleasure, whether through the objectification of the celebrities whose photos hang in teenage girls' rooms or in Internet teen chat rooms where persons of the opposite sex meet,

construct themselves, and experiment with interactions.

Because the "pure" relationship is not anchored in anything beyond itself, Giddens argues, voluntary commitment plays a central role, and therefore the ability to trust the other becomes central to its continuance. Intimate communication which validates and develops the self, therefore, seems to be an integral goal of the "pure" relationship of which Giddens writes. The other side of the coin, then, is that teens - much like adults - seek to avoid emotional risk because they see it as a potential threat to the *self*. Some of the teens and their parents with whom I spoke noticed this tendency among teens to avoid emotional investment. Paula Hall, who is the mother of Michael (a case study and peer discussion group leader), contrasted this with her own generation's proclivity to seek "deep" relationships:

> Paula: My kids, this generation, they're very surface. They don't get deep.
> Michael (15 years old): See, we're mellow.
> Paula: Back in the '60s, everybody got real deep. Even though they were into the free love and all that, they got into it real deeply. But this gerenation is more of, 'don't make me go that far down into the situation.'
> Lynn (to teens): Do you think that's true?
> Tammy (19 years old): I think it's true, and I think that's because of the way things are now. I feel that even with me, and a lot of my friends, there's so much violence, it's like, I don't know anybody that hasn't lost a good friend or brother or cousin or whatever to violence, so it's like, we don't try to get real deep in stuff. Everything's like, 'whatever.' If something happens it's like, 'I don't care.'
> Michael: And like, when a movie that comes out like that's all silliness, it kinda lets you escape from all the seriousness and stuff like that.
> Tammy: But even, most serious movies, that are supposed to be real violent, my mom gets mad at us, 'cause she gets all sentimental about it, 'oh, it's so terrible,' and we're all like laughing at her. She always says we're morbid. And I'm like, well, you can't really get that into it, 'cause with the everyday thing, you'll just go crazy. You gotta kinda go with the flow.

Lynn: So, is there something that gives you hope that things will be different sometime?
Tammy: I don't know. (chuckles) Well.
Michael: Well, hope's just one day at a time.

In this exchange, the older teen explains the motivation for the avoidance of emotional risk, or "depth" in relationships, in terms of violence and the potential for loss. In fact, there is so much loss in "the everyday thing" for her that she explains her own desensitization to media portrayals of violence as a part of her position of self-defense. In doing so, she demonstrates the way in which self-construction and self-preservation play an important role in determining personal relationships and in "reading" the cultural and mediated texts of relationships, as well.

Teens have long been conscious that their "task" is to find their place in society, as they and their families are influenced by the Erickson-inspired discourse of developmental psychology. They know that they are to develop a sense of identity during their teen years and thus with varying levels of consciousness they experiment with practices and explanations for who they are and why they do what they do. Yet the rise of the "risk" society, coupled with increased autonomy and the politicization of identities (racial/ethnic identies in particular) have taught them to be extremely aware of their narrative constructions. Giddens (1991) points out that in a society ever-conscious of changes and risks, it becomes increasingly important to "keep a particular narrative going (p. 54)." Bringing Giddens and Hall together, therefore, I argue that this is done as the individual continuously reflects on his or her experiences, incorporating new discursive strategies while discarding others. It is a reflexive process not unrelated to what Lasch (1978) calls the "therapeutic

sensibility" prevalent in the U.S. Individuals, Lasch argues, "hunger not for personal salvation, let alone for the restoration of an earlier golden age, but for the feeling, the momentary illusion, of personal well-being, health, and psychic security (p.7)." Teens' constructions of identity, and indeed their senses of their roles as individuals within the larger society, seem to be shaped by this sensibility.

This rise in reflexivity is related to the second aspect of teen life I believe shapes the contexts for religious identity-construction, which I have called the flattening of various social hierarchies. In many ways these trends continue the upheavals begun in the social revolutions of the 1960s, when the parents of today's teens were in the formative years of early adolescence and early adulthood. In chapter six I noted the perceptions, voiced by parents of teens, that the relationships between parents and teens are different today than they were when the parents themselves were teens. They report an ambiguity regarding issues of authority which extend from uncertainties regarding punishment practices to their own pleasure in participating *in*, rather than simply *observing* (as their parents did), their teenage children's activities. I suggested that this may be somewhat related to the rise in single parenthood and the desire for "meaningful" relationships between parents and teens.

This change in family structure also relates to the sense that teens' perceptions of social arrangements are, in general, perhaps less hierarchical than in their parents' generation. While some parents may have grown up surrounded by a majority of what one parent termed the "normal" families of two parents and two children, this generation of

teens has experienced a greater diversity in family structure and is much more hesistant to judge certain family structures as preferred, or certainly more "normal," than others. Similarly, today's teens have grown up in an environment of increasing racial/ethnic diversity and within the politicized environments of a heightened awareness of injustice and difference. While I have demonstrated that the amount of activism among teens in this area may be limited, it is clear that all teens know that they *should* be tolerant of difference, although understandings and practices of that tolerance vary. Whether with celebration or reluctance, today's teens have experienced the crushing of former assumptions which placed certain races above others biologically and culturally. Even as their parents see themselves as more tolerant than their own parents, teens see themselves (and their parents see them, as well) as more tolerant than their parents. Hence hierarchies in this social arena continue a trend of being undermined, as well.

The parents of teens also report a flattening of hierarchies between the choices of social activities in which their teens take part. With the urbanization and suburbanization of the U.S., schools have become larger and their students more anonymous than was the case when the teens' parents were growing up. Thus as schools no longer serve as extensions of the homogeneous community, they are also less frequently the sole location of adolescent activity and identification. Each of the case study teens reviewed here, for example, participates in activities in communities at some distance from their own home and school, whether it be in a religious organization, a community center, or even amidst fellow drag racing enthusiasts. Thus their school friends have social networks in

various locations, and individual adolescents become aware of the myriad of possible alternative sources of activities and identifications. There is thus no longer a clearly "better" alternative for teens or a hierarchy of choices; only a sense that as each teen chooses identifications with social networks based on her needs, she finds the best alternatives *for her.*

What might be the implications of this rise in reflexivity and flattening of hierarchies as teens come to their mediated experiences? Further, what might these imply regarding their construction of religious identity? To address these issues, we must first turn to the second analytical task of this dissertation: an exploration of the discourses of religion that are available in mediated texts and echoed throughout the culture.

The Discourses of Religion

To explore some of the discourses of religion available in U.S. culture, I began with the mediated text, *Touched by an Angel.* Because the year of this study saw the introduction of the largest number of prime-time television programs dealing with some aspect of religion in recent history (or perhaps ever), there were a number of other possible programs to select, and I was able to use some of these in questions of comparison to elicit more information on a subject's positioning *vis a vis* these texts. Still, parents occasionally asked why I had selected CBS's *Touched by an Angel* as opposed to the WB Network's *7th Heaven,* as the latter is arguably more popular with teens. Moreover, the WB Network programs to the teen audience whereas CBS struggles against its stereotype as the "gray" network. As I noted earlier, I chose the program in part because of

its widespread popularity which does in fact cut across generations, but also because I was interested in discussing issues of religion and believed that most people would identify the program as at least broadly having something to do with religion or spirituality. In *Touched,* for instance, the angels are central to the plot of each story. In contrast, while *7th Heaven* features a pastor and his family, the central plots of this program usually have to do with the family. Occasionally there are references to the religious foundations for moral choices in the latter, but I believe that *Touched* is perhaps more polysemic, or more "open," as a number of respondents stated, to various discursive positions of religion. Moreover, while *7th Heaven'* s all-white cast occasionally deals with issues of race, *Touched by an Angel* has an African American in a leading role and in the episode I utilized, African Americans in several other key roles, as well. I think that future research comparing the reception of the two programs may yield further insights into the role of overtly religious symbols (such as the clerical collar and church building featured in *7th Heaven)* and the more diffuse religious symbols of *Touched'* s angels, and I hope to further explore this.

As I noted in the previous chapter, I recognize that while *Touched by an Angel* is open to several discursive strategies of religion, the possibilities are not exhaustive, and thus my analysis does not address all of the possible discourses of religion available in the U.S. today. Still, I believe it was quite useful to focus on the single, popular text, as this provided some means for comparison between the various conversations that emerged in response to it and provided some clues (and limits) to the religious discourses I wanted to, and could, study. My early interviews in

this project had included a series of general discussions on religion, values, identity, and media consumption practices with persons from very different backgrounds which, while helpful, often generated conversations that ranged so broadly that they were difficult to compare. By employing a single program I was able to explore not only which disourses from the program were salient among those I interviewed, but I also was provided with data that allowed me to demonstrate how at least some of the salient and popular discourses on religion in the U.S. are related to the historically powerful and hegemonic assumptions of Protestantism.

I pointed out in chapter six that I discovered three different yet overlapping discursive approaches to religion among the parents' groups: one that emphasized morality, another that focused on religion as a source of ultimate meaning, and a third that approached religion as one of many possible "compartments" in life that may or may not have anything to do with the other parts. In chapter seven, I then named seven possible discursive approaches to religion I observed in my conversations with teens, both around the specific program and in other contexts, as well. I detailed an *institutional* discourse which tended to equate religion with the institutions of organized religion, a *revelatory* discourse which sought God's message of love, equality and morality in such places as a mediated text or even a mother's conversation, and a *sentimental* discourse which saw religion as something vaguely positive that created good feelings, but did not have much relevance beyond these feelings in everyday life. I also noted the discourse of *morality* which I believe was encoded in the program itself and also articulated by several individuals. Additionally, I briefly discussed the *mystical* discourse I had highlighted in chapters two

and four (with Joe and Amber's Star Trek/paranormal narrative and Jodie's X-Files/aliens narrative) and noted the fact that other studies in religion had revealed an *evangelical* and *social justice* discourse, neither of which emerged among teens in my study. Participants in the third peer-led discussion group, who saw themselves as 'othered' by mainstream culture due to their racial/ethnic and (to a lesser extent) religious backgrounds voiced a discourse of *difference* in relation to the program and to each other; rather than responding to one of the discursive strategies embedded in the text, members of the group employed contrasting discursive strategies to assert their religious and racial/ethnic identities.

I thus argued in chapter seven that the program *Touched by an Angel* allows for viewers to interpret the program differently according to the discursive position they articulate in their own identity narratives. It is, of course, possible to see overlaps between the parents' conversations and the teens. The conversations on morality in 15-year-old Michael's group of conservative Protestants and nonreligious members held some similarity to those emerging in both the conservative Roman Catholic Ortiz group and among the Protestant, Roman Catholic, and nonreligious members of Cathy Swenson's groups, for example. Several individuals in these groups approached religion as fundamentally and primarily tied to questions of morals, as I described in chapters six and seven. There are also similarities in the ways in which both one set of parents in the moderate Protestant Gray's group and Jake, the teen with no religious background, seemed to compartmentalize religion and relate it to positive, "sentimental" feelings. Additionally, while some members of 15-year-old

Elizabeth's group looked favorably on organized religion and seemed to see all religion as encompassed in its institions, there were a number of other adults and teens who similarly equated religion with its institutions, which in some cases became a grounds for the dismissal of all considerations of religion.

It is also interesting to note that the discussion of religion as a source of ultimate meaning only occurred in the Gray's group, and then when it was suggested as such by the pastor in that group. This may simply suggest that such an approach is too abstract or at least difficult to articulate in a narrative mediated form such as *Touched by an Angel* and thus conversations centered on the mediated text did not reveal this discursive approach. It may also suggest that this is a less salient approach to contemporary religion. As so much of orthodox Protestant and Catholic theological writings, based in the works of such persons as Tillich and Schleiermacher. emphasize religion as an experience of "ultimate meaning," further research in this area could yield interesting results that may provide clues on why some persons find the messages of organized religion (particularly of the moderate-to-liberal variety) less than salient.

It is interesting to note the similarities in discursive strategies that cross the lines of various institutional affiliations (or lack thereof). This demonstrates that the approaches to religion articulated in the religious identity narratives of individuals may exist apart from the affiliations individuals hold in religious institutions. This then reaffirms the assertion made in the first chapter regarding the decline of the authority of religious institutions and the rise of personal autonomy. Such a situation, therefore, reiterates the importance of the mediated realm (as opposed to

primarily or solely the realm of the religious institution) in the individual's construction of religious identity.

In chapter seven, I also discussed three overarching themes that emerged in the discussions with teens: the concept of a universal God, the importance of tolerance for differing religions, and the assumption that religion is primarily an individual and private choice. While the latter two had been mentioned in chapter six, the notion of a universal God was not mentioned among the parents. As tolerance seemed to be a particularly salient issue for all of the groups, I then reviewed how this latter belief related to the so-called "openness" of *Touched by an Angel.* I concluded there that the while the program allowed individuals with at least some marginal affiliation with Christianity to believe that the program is "open" to those of "other" faiths, its Christian-centric themes (clearly identified as such by those who actually do embrace other faiths) actually provided evidence of the continuing, if diffuse, power of Protestant hegemony in the contemporary U.S. This demonstrates that while Protestant Christianity's hegemony may be increasingly undermined, the process is occurring dialectically. Thus it is not surprising to see contradictions and fissures in the discourses of religion as they emerge both in mediated texts and in the religious identity narratives that relate to those texts.

As I noted in the previous two chapters, many teens seem to recognize the importance of religious tolerance and, like their parents, object to the idea that one approach to religion might be superior to others. Thus I argue that regardless of how tolerance is approached in practice, it is important to see the acceptance of the *notion* of tolerance for other

religions as another example of a flattening of social hierarchies. This, too, has implications for the process of religious identity-construction among teens, as I will discuss in the next and last section.

Implications for the Interpretation of Religious Symbols

The flattening of social hierarchies reported by teens and their parents is, of course, in the realm of *perception* to a large extent. I equate this with the trends of postmodernism, although I share Huyssens's (1984) hesitation in describing the postmodern as a wholly transformative condition:

> What appears on one level as the latest fad, advertising pitch and hollow spectacle is part of a slowly emerging cultural transformation in Western societies, a *change in sensibility* for which the term 'post-modern' is actually, at least for now, wholly adequate. The nature and depth of that transformation are debatable, but transformation it is. I don't want to be misunderstood as claiming that there is a wholesale paradigm shift of the cultural, social, and economic orders; any such claim clearly would be overblown. But in an important sector of our culture there is a noticeable shift in *sensibility, practices and discourse formations* which distinguishes a post-modern set of assumptions, experiences and propositions from that of a preceding period (cited in Harvey, 1990, p. 39, italics mine).

One could also see the similarities between what I am calling the flattening of hierarchies and the resistance to meta-narratives articulated by Foucault (1972). However, while I have highlighted these social changes and have also focused on the concept and role of discourses, I do not intend to argue that all power is now wholly lodged in discursive practices. While I have noted, for instance, that the roles of the religious institutions in religious identity-construction are far from determinative, I have not completely located the constructive power within individual

autonomy as an "active audience researcher might, or in structured "interpretive communities" (Fish, 1980), or even in discourses developed in a particular bureaucratic and somewhat apolitical forum, as implied in Foucault's argument. Rather, the discourses of religion I have discussed should be seen in light of their historical development in *relation* to religious institutions and practices, but also in relation to a more diffuse "popular religiosity" as well as in relation to the various political and economic power systems they very indirectly continue to hold in place.

Yet clearly, I am not wholeheartedly rejecting postmodern theory. The flattening of hierarchies I note, for example, may be seen as connected to the ephemerality of the postmodern condition which, within the context of a "risk" society, encourages young people to seek security wherever they can find it. As Harvey (1990) argues:

> Deeper questions of meaning and interpretation also arise. The greater the ephemerality, the more pressing the need to discover or manufacture some kind of eternal truth that might lie therein. The religious revival that has become much stronger since the late sixties, and the search for authenticity and authority in politics (with all of those charismatic and 'protean' individuals with their Nietzschian 'will to power') are cases in point. The revival of interest in basic institutions (such as family and community), and the search for historical roots are all signs of a search for more secure moorings and longer-lasting values in a shifting world (p. 292).

One can see points of connection with the earlier comments on lack of "depth" in relationships by 15-year-old Michael Pearson's family, and also with the data reviewed earlier, particularly in the parents' tendency to equate a need for security and "hope" among their young people with these "basic institutions" of family, religion, and community.

There are also other connections with postmodern theory implied

in this study. For instance, the pragmatic approach to symbols I discussed in chapter four may also be considered related to the understanding of the sign as a commodity with a "use-value" within late capitalism, as Jameson (1991) has pointed out in his critique of postmodernism. This, in fact, is perhaps a key aspect of the teen-media relationship that must be considered in the question of religious identity-construction, and thus it constitutes a primary finding of this project. In addition to the flattening of social hierarchies, I argue that teens experience a *flattening of religious symbols.* As I noted in chapter four, Fowler (1978, 1981), the chief theorist in stagewise religious development, argues that for most teens, the signs which refer to the belief systems they adhere to are held to be sacred; thus when the sign is defamed, the belief system is, as well. Yet as noted in many interviews and observations with teens, very few of them seem to adhere to such a strong connection between religious symbols and their referents, regardless of the depth, coherence, or ideological commitment of the teens' religious beliefs. Religious symbols, like all others in a postmodern framework, are not considered meaningful in and of themselves. They must be put into a context that *makes* them meaningful; they must be *made useful* in this meaning-making process. The popular discourses of religion emerging in the media may aid in this process, or at least echo it. This is why, then, such seemingly contradictory mediated symbols as those emerging on science fiction programs such as *Star Trek* or *The X-Files* may be just as useful in religious identity-construction as an "overtly religious" program such as *Touched by an Angel.* Distinctions may be made by teens, but both types of programs still seen to exist within, as Elizabeth terms it, a broad realm of "beliefs." Teens

accept that programs dealing with "beliefs" may be made useful by different teens, or even by themselves at different times depending on the circumstances. In fact, those religious symbols that seem more closely tied to the historical institutions of the Christian religion - such as the clerical collar and the church building - seem to be *less* useful. If one aspect of the appeal of more "general" images such as angels is that they may be considered "open" - whether in the sense that persons of other faith traditions might through them be convinced of Christianity or simply find them meaningful within their own systems - then these symbols more directly tied to the institutions of Christianity are *less useful,* in large part because of the emergent *sensibility of tolerance.*

One of the interesting consequences of this "openness" of religious symbols is that the original context of these religious symbols is of less importance, and thus the meanings of the symbol become increasingly diffuse. I observed that in the fact that a number of teens saw little distinction between "angels" and "ghosts," as discussed in relation to 17-year-old and nonreligious Jake's approach to *Touched by an Angel.* The short-lived Fall 1997 ABC sitcom *Teen Angel,* featuring a deceased boy whose ghost(?) becomes a "guardian angel" to his best friend, articulated exactly this position. Yet teens are not overly concerned with the referents unless they become *useful* to them; thus the vagueness of definitions is not worrisome; in fact, it is hardly noteworthy. Further research is needed on this blurring of boundaries between the reception (and possibly *regeneration)* of those symbols with some reference to historical religious institutions (such as angels) and those emerging out of more "popular" contexts (such as ghosts). The mediated environment, which presents

images of both, provides a rich resource for the further analysis of the interpretation of these symbols through reception studies such as this.

Thus I conclude with the question once posed to me by Klaus Bruhn Jensen (1996): "*Can* there be a postmodern religion?" In this dissertation, I have argued that religion, and religious identity, are embedded in and cannot be understood apart from the environment in which they exist. Furthermore, this process of religious identity-construction occurs within a largely mediated environment in which both subjectivities and discourses are created and maintained within a context of postmodern sensibilities. Combining each of these factors with the teens' approach to religion as that which only exists as it is constructed into useful form, then we might argue that the *only* possible religious identity for contemporary teens today is that of a postmodern religion.

In this dissertation I have attempted to describe the interaction between sites of determination in popular media texts and their audiences in the identity-construction process, focusing on patterns of interpretation that I have related to larger cultural discourses of religion. I have challenged the essentialist notions of a "religious" audience and the monolithic assumptions of the mediated realm by demonstrating how young people negotiate their religious identities within the contexts of a "risk" society, postmodern sensibilities of tolerance and the flattening of numerous social hierarchies, and a pragmatic approach to the landscape of flattened religious symbols. What will this mean for the future of the media/religious identity/youth nexus that has been presented here? Perhaps we will have to tune in to programs like*The X-Files* (and their audiences) to find out.

References

Abu-Lughod, L. (1993). **Writing women's worlds: Bedouin stories.** Berkeley: University of California Press.

Adorno, T. (1991). How to look at television. **The Culture Industry.** London: Routledge.

Albanese, C. (1993). Fisher kings and public places: The old new age in the 1990s. **Annals of the American academy of religion** 527 (May), 131-143.

Alexander, A., Ryan, M.S., & Munoz, P. (1984). Creating a learning context: Investigations on the interaction of siblings during television viewing. **Critical studies in mass communication,** 1, 345-364.

Amesley, C. (1989). How to watch *Star Trek.* **Cultural Studies** 3, 323-39.

Ang, I. (1985). **Watching *Dallas:* Soap opera and the melodramatic imagination.** London: Methuen.

Ang, I. (1989). Wanted: Audiences. On the politics of empirical audience studies. In E. Seiter, H. Borchers, G. Kreutzner, and E. Warth (Eds.), **Remote Control: Television Audiences and Cultural Power.** Routledge: London.

Ang, I. (1996). **Living room wars: Rethinking media audiences for a postmodern world.** London: Routledge.

Arnold, M. (1954). **Poetry and prose.** London: Rupert Hart Davis, p. 343. Cited in Storey, J. (1993). **Cultural theory and popular culture.** Athens: University of Georgia Press.

Arnold, M. (1960). **Culture and anarchy.** London: Cambridge University.

Atkin, C.K. (1990). Television advertising and the socialization to consumer roles. In D. Pearl, L. Bouthilet, & J. Lazar (Eds.), **Television and behavior: Ten years of scientific progress and implications for the eighties** (Vol. 2, pp. 191-200). Rockville, MD: National Institute of Mental Health.

Arnold, C. (1997). Twilight Zone (*Touched by an Angel* star Roma

Downey gets used to celebrityhood). **People Weekly** 47(23), p. 140.

Appadurai, A. (1991). Global ethnoscapes: Notes and queries for a transnational anthropology. In R. Fox, (Ed.), **Recapturing anthropology: Working in the present.** Santa Fe: SAR Press.

Babcock, B. (1980). Reflexivity: Definitions and discriminations. **Semiotica** 30 (1/2), 1-14.

Bachman, J.G., Johnston, L.D., & O'Malley, P.M. (1993). **Monitoring the future: Questionnaire responses from the nation's high school seniors, 1992.** Ann Arbor, MI: Survey research center, Institute for social research, the University of Michigan. In B.A. Chadwick & T.B. Heaton (Eds.)., **Statistical handbook on adolescents in America.** Phoenix: Oryx Press.

Balmer, R. (1989). **Mine eyes have seen the glory: A journey into the subculture of evangelicalism.** New York: Oxford University Press.

Bandura, A. (1977). **Social learning theory.** Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., Ross, D., & Ross, S.A. (1963). Imitation of film-mediated aggressive models. **Journal of abnormal and social psychology,** 66, 3-11.

Baudrillard, J. (1981). **For a critique of the political economy of the sign.** St. Louis: Telos Press.

Behar, R. (1993). **Translated woman: Crossing the border with Esperanza's story.** Boston: Beacon Press.

Bellah, R., Madsen, R., Sullivan, W., Swidler, A., and Tipton, S. (1985). **Habits of the heart: Individualism and commitment in American life.** New York: Harper and Row.

Benedict, R. (1934/1961). **Patterns of culture.** Boston: Houghton Mifflin Company.

Benson, P.L. (1993). **The troubled journey: A portrait of 6th-12th grade youth.** Minneapolis, MN: Search Institute.

Berger, P. (1969). **The Sacred Canopy: Elements of a sociological theory of religion** (1st ed). Garden City, N.Y.: Doubleday.

Berger, P. and Luckmann, T. (1967). **The social construction of reality.** Harmondworth: Penguin.

Berkowitz, L. (1964). The effects of observing violence. **Scientific American** 21, 35-41.

Berkowitz, L. & Rawlings, E. (1963). Film violence and cue properties of available targets. **Journal of Personality and Social Psychology** 3, 525-530.

Bezilla, R. (Ed.) (1993). **America's youth in the 1990s.** Princeton, NJ: The George H. Gallup International Institute.

Bibby, R. (1987). **Fragmented Gods.** Toronto: Irwin.

Bird, E. (1992). Travels in nowhere land: Ethnography and the "impossible" audience. **Critical studies in mass communication** 9(3), 250-260.

Bird, E. (1995). Personal conversation during the annual meeting of the International Communication Association, Albuquerque, New Mexico.

Blumer, H. (1933). **Movies and conduct.** New York: Macmillan.

Blumer, H. (1969). **Symbolic interactionism.** Englewood Cliffs, NJ: Prentice-Hall.

Bourdieu, P. (1977). **Outline of a theory of practice.** Cambridge: Cambridge University Press.

Brill, S. (1997). Martha's Angels: A touch of the divine in prime time. **Christianity Today** 39(13), 65.

Broch, H. (1968). Evil in the value system of art. In G. Dorfles (Ed.), **Kitsch: The world of bad taste.** New York: Bell Pub. Co. Cited in McDannell, C. (1995). **Material Christianity.** New Haven: Yale University Press.

Brown, J.D., Klein, J. & Walsh-Childers, K. (1990, July). Risky Behavior and media use among adolescents. Paper presented at the meeting of the International Communication Association, Dublin, Ireland. Cited in Brown, J. (1993). Theoretical Overview. In B.S. Greenberg, J.D. Brown, and N.L. Buerkel-Rothfuss (Eds.), **Media, sex and the adolescent.** (pp. 19-25). Cresskill, NJ: Hampton Press.

Buckingham, P. (1993). Introduction: Young people and the media. In

P. Buckingham (Ed.), **Reading audiences: Young people and the media.** Manchester and New York: Manchester University Press.

CBS (Producer). (1997). **60 Minutes.** November 9, New York.

Carey, J. (1989). **Communication as culture: Essays on media and society.** London: Routledge.

Carter, S. (1993). **The culture of disbelief.** New York: Basic books.

Casanova, J. (1994). **Public religions in the modern world.** Chicago: The University of Chicago Press.

Chadwick, B.A. & Heaton, T.B. (Eds.)., **Statistical handbook on adolescents in America.** Phoenix: Oryx Press.

Cirksena, L. and Cuklanz, L. (1992). Male is to Female as ___ is to ___: A Guided Tour of Five Feminist Frameworks for Communication Studies. In L. Rakow (ed.), **Women Making Meaning.** (pp. 18-44). New York: Routledge.

Clark, L.S. and Hoover, S. (1997). Controversy and cultural symbolism: Press relations and the formation of public discourse in the case of the RE-Imagining event. **Critical studies in mass communication** 14, 310-331.

Clark, L. S. (1996a). Media, meaning and identity-construction in the postmodern context: The case of U.S. adolescents and their families. Presented at the Network of Qualitative Television Audience Research panel during the annual meeting of the International Association of Mass Communication Researchers, Sydney (August).

Clark, L.S. (1996b). Media, meaning and religious identity: A case study. Paper presented at the annual mmeting of the Society for the Scientific Study of Religion, St. Louis (November).

Clark, L.S. (1996c). Media, meaning and the religious self: An ethnographic case study. Paper presented at a colloquium presentation at the School of Journalism and Mass Communication, Boulder (November).

Clifford, J. (1986). Introduction. In Clifford, J. and Marcus, G. (Eds.), **Writing culture: The poetics and politics of ethnography.** Berkeley: University of California Press.

Clifford, J. and Marcus, G. (Eds.). (1986). **Writing culture: The poetics and politics of ethnography**. Berkeley: University of California Press,

Coffield, F. (1987). From the celebration to the marginalization of youth. In G. Cohen (Ed.), **Social change and the life course**. London: Tavistock Publications.

Collins, D. (1980). **Sartre as biographer**. Cambridge, Mass.: Harvard University Press.

Comstock, G., Chaffee, S., Katzman, N., McCombs, M., and Roberts, D. (1978). **Television and human behavior**. New York: Columbia University Press.

Cooley, C.H. (1902/1956). **Human nature and the social order**. New York: Scribner's/Free Press. Cited in Denzin, N. (1992). **Symbolic interactionism and cultural studies**. Oxford: Blackwell.

Corner, J., Richardson, K. and Fenton, N. (1990). **Nuclear reactions: Form and response in 'public issue' television**. London: John Libbey. Cited in Moores, S. (1995). **Interpreting audiences: The ethnography of media consumption**. London: Sage.

Coupland, D. (1991). **Generation X**. New York: St. Martin's Press.

Coupland, D. (1994). **Life after God**. New York: Pocket books.

Coward, R., and Ellis, J. (1977). **Language and materialism**. London: Routledge and Kegan Paul. Cited in Sampson, E. (1989). Deconstructing the Self. In J. Shotter and K. Gergen (Eds.), **Texts of identity** (pp. 1-19). London: Sage.

CNN/USA Today/NSF/Gallup Poll. (1997). National survey of teenagers, March 20-27, 1997. (Available from The Gallup Organization, 47 Hulfish Street, Suite 200, Princeton, New Jersey 08542).

Craig, S. (Ed.). (1992). **Men, masculinity and the media**. Newbury Park, CA: Sage.

Csikszentmihalyi, M. & Larson, R. (1984). **Being adolescent: Conflict and growth in the teenage years**. New York: Basic books.

Dayan, D. and Katz, E. (1992). **Media events: The live broadcasting of history**. Cambridge: Harvard University Press.

deLauretis, T. (1984). **Alice doesn't: Feminsm, semiotics, cinema.** Bloomington: Indiana University Press.

Demott, B. (1990). **The imperial middle: Why Americans can't think straight about class.** New Haven: Yale University Press.

Denzin, N. (1992). **Symbolic interactionism and cultural studies.** Oxford: Blackwell.

Derrida, J. (1978). **Writing and difference.** Chicago: University of Chicago Press.

Douglas, M. (1970). **Natural symbols: Explorations in cosmology.** New York: Pantheon Books/ Random House.

Dykstra, C. (1989).Editorial. **Theology Today 46** (July), 127. Cited in Roof, W.C. (1993). **A Generation of seekers: The spiritual journeys of the baby boom generation.** San Francisco: Harper.

Elder, G.H. (1975). Adolescence in the life cycle: An introduction. In S.E. Dragastin and G.H. Elder (Eds.), **Adolescence in the life cycle: Psychological change and social context** (pp. 1-22). New York: John Willey and Sons.

Eller, C. (1993). **Living in the lap of the goddess: The feminist spirituality movement in America.** New York: Crossroad.

England, K. (1994). Getting personal: Reflexivity, positionality, and feminist reseach. **The professional geographer** 46(1), 80-89.

Ethan Allen, Inc. (1986). **The Ethan Allan Report: The status and future of the American family.** Danbury, CT: Ethan Allen, Inc. In Chadwick, B.A. & Heaton, T.B. (Eds.)., **Statistical handbook on adolescents in America.** Phoenix: Oryx Press, p. 226.

Erikson, E. (1968). **Identity, youth, and crisis** (1st ed.). New York: W.W. Norton.

Erikson, E. (Ed.). (1965). **The challenge of youth.** Garden City, N.Y.: Doubleday.

Ewen, S. (1988). **All Consuming Images: The politics of style in contemporary culture.** New York: Basic books, 1988.

Fallers Sullivan, W. (1995). Diss-ing religion: Is religion trivialized in

American public discourse? **The Journal of Religion** 75(1), 69-79.

Ferguson, R. Gever, M., Minh-ha, T., and West, C. (Eds.) (1990). **Out there: Marginalization and contemporary cultures.** Cambridge, MA: MIT Press.

Fine, G., Mortimer, J., and Roberts, D. (1990). Leisure, work, and the mass media. In Feldman, S. and Elliott, G. (Eds.). **At the threshold: The developing adolescent** (pp. 225-252). Cambridge: Harvard University Press.

Fish, S. (1980). **Is there a text in this class? The authority of interpretive communities.** Cambridge, MA: Harvard University Press.

Fiske, J. (1987). **Television culture.** London and New York: Routledge.

Fiske, J. (1989). **Understanding popular culture.** London: Unwin Hyman.

Flitterman-Lewis, S. (1992). Psychoanalysis, film theory and television. In Allen, R. (Ed.), **Channels of discourse, reassembled** (pp. 203-246). Chapel Hill: University of North Carolina Press.

Foucault, M. (1972). **Power/knowledge.** New York.

Foucault, M. (1988). **The care of the self.** Harmondsworth: Penguin.

Foucault, M. (1987). **The use of pleasure.** Harmondsworth: Penguin.

Fowler, J. (1981). **Stages of faith: The psychology of human development and the quest for meaning.** San Francisco: Harper and Row.

Fowler, J. & Keen, S. (1978). **Life maps: Conversations on the journey of faith.** Waco, TX: Word books.

Fornas, J., Lindberg, U., and Sernhede, O. (1995). **In garageland: rock, youth, and modernity.** Trans. by Jan Teeland. London, New York: Routledge.

Freud, S. (1961). **The future of an illusion.** (J. Strachey, Trans. and Ed.). New York: W.W. Norton & Company.

Frith, S. (1983). **Sound effects: Youth, leisure and the politics of rock.** London: Constable.

Funk, J. (1992). Video games: Benign or malignant? **Journal of**

developmental and behavioral pediatrics 13, **53-54.**

The George H. Gallup International Institute. (1992). **The religious life of young Americans**. Princeton, NJ: The George H. Gallup International Institute.

The George G. Gallup International Institute. (1991). **The Gallup survey on teenage suicide**. Princeton, NJ: The George H. Gallup International Institute.

Garnham, N. (1995). Political economy and cultural studies: Reconciliation or divorce? **Critical studies in mass communication** 12 (3), 62-71.

Gates, H.L. (1992). **Loose canons: Notes on the culture wars**. New York: Oxford University Press.

Geertz, C. (1973). **The interpretation of cultures**. New York: Basic Books.

Geertz, C. (1968). **Islam observed: Religious development in Morocco and Indonesia**. Chicago: University of Chicago Press.

Geertz, C. (1979). From the natives' point of view: On the nature of anthropological understanding. In Rabinow, P. and Sullivan, W.M. (Eds.), **Interpretive social science**. Berkeley, University of California Press, 225-241, p. 229. Cited in Sampson, E. (1989). The deconstruction of the self. In Shotter, J. and Gergen, K. (Eds.), **Texts of Identity** (pp. 1-19). London: Sage.

Gergen, K. (1991). **The saturated self: Dilemmas of identity in contemporary life**. New York: Basic books.

Giddens, A. (1984). **The constitution of society**. Berkeley: University of California Press.

Giddens, A. (1991). **Modernity and self-identity: Self and society in the late modern age**. Stanford: Stanford University Press.

Gillespie, M. (1993). The *Mahabharata:* from Sanskrit to sacred soap. A case study of the reception of two contemporary televisual versions. In Buckingham, D. (Ed.), **Reading audiences: Young people and the media.** New York: Manchester University Press, 48-73.

Gilligan, C. (1982). **In a different voice: Psychological theory and women's development.** Cambridge, Mass. : Harvard University Press.

Gilligan, C., Lyons, N., & Hanmer, T. (Eds.). (1990). **Making connections: The relational worlds of adolescent girls at Emma Willard School**. Cambridge, MA: Harvard University Press.

Glaser, B. & Strauss, A. (1967). **The discovery of grounded theory: Strategies for qualitative research**. New York: Aldine de Gruyter.

Golding, P. & Murdock, G. (1981). Ideology and the mass media: The question of determination. In M. Barrett, et. al. (Eds.), **Ideology and cultural production**. New York: St. Martin's Press.

Gramsci, A. (1971). **Selections from the prison notebooks** (Q. Hoare and G. N. Smith, Trans. and Eds.). New York: International Publishers.

Greenberg, B.S., Brown, J.D., & Buerkel-Rothfuss, N. (1993). **Media, sex and the adolescent**. Cresskill, NJ: Hampton.

Greenfield, P.M. Bruzzone, L., Koyamatsu, K., Satuloff, W., Nixon, K., Brodie, M., & Kingsdale, D. (1987). What is rock music doing to the minds of our youth? A first experimental look at the effects of rock music lyrics and music videos. **Journal of Early Adolescence** 7, 315-329.

Griffiths, V. (1995). **Adolescent girls and their friends : A feminist ethnography**. Aldershot: Avebury; Brookfield, USA: Ashgate.

Grodin, D. and Lindlof, T. (1996). (Eds.), **Constructing the self in a mediated world**. Thousand Oaks, CA: Sage.

Grossberg, L. (1987). Critical theory and the politics of empirical research. In M. Gurevitch and M.R. Levy (Eds.), **Mass Communication Review Yearbook** 6, 86-106.

Guba, E. & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. Lincoln (Eds.), **Handbook of qualitative research**. Thousand Oaks, CA: Sage.

The Alan Guttmacher Institute. (1994). **Sex and America's teenagers**. New York: The Alan Guttmacher Institute.

Habermas, J. (1992). **The structural transformation of the public sphere**. Transl. by T. Burger. Cambridge, MA: MIT Press.

Habermas, J. (1979). **Communication and the evolution of society**. Boston: Beacon Press.

Hall, G.S. (1916). **Adolescence.** New York: Appleton.

Hall, S. (1980). Encoding/decoding. In S. Hall, D. Hobson, A. Lowe, and P. Willis (Eds.), **Culture, media, language.** London: Hutchinson, 128-138.

Hall, S. (1981). Notes on deconstructing the popular. In Samuel, R. (Ed.) **People's history and socialist theory.** London: Routledge.

Hall, S. (1982). The Rediscovery of 'ideology:' The return of the repressed in media studies. In M. Gurevitch, T. Bennett, J. Curran, and J. Woollacott, (Eds.), **Culture, Society, and the Media.** (pp. 56-90). London: Methuen.

Hall, S. (1989). Cultural identity and cinematic representation. **Framework** 36, 68-81.

Hall, S. (1996). Introduction: Who needs 'identity'? In S. Hall and P. duGay (Eds.), **Questions of Cultural Identity** (pp. 1-17). London: Sage.

Hall, S. & Whannel, P. (1964). **The Popular Arts.** London: Hutchinson.

Hall, S. & Jefferson, T. (Eds.) (1978). **Resistance through rituals.** London: Methuen.

Hall, S. & du Gay, P. (Eds.). (1996). **Questions of cultural identity.** London: Sage.

Hammond, P. (1992). **Religion and personal autonomy : The third disestablishment in America.** Columbia, SC: University of South Carolina Press.

Haraway, D. (1985). Cyborg manifesto and fractured identities. **Socialist Review** 80, reprinted in Lemert, C. (Ed.) (1993. **Social theory: The classic and multicultural readings.** Boulder: Westview press, 597-604.

Peter D. Hart Research Associates. (1996). **Children/parents: Television in the home.** (Tech. Rep. No. 1). Philadelphia, PA: The University of Pennsylvania, Annenberg Public Policy Center.

Harvey, D. (1990). **The condition of postmodernity.** Cambridge MA and Oxford: Blackwell.

Hass, N. (1998, February 22). A TV generation is seeing beyond color. **The New York Times**, section 2, pp. 1, 38.

Hebdige, D. (1979). **Subculture: The meaning of style**. London: Methuen.

Hebdige, D. (1976). Reggae, rastas and rudies. In S. Hall and T. Jefferson (Eds.), **Resistance through Rituals: Youth subcultures in post-war Britain**. London: Center for Contemporary Cultural Studies.

Hirst, P. (1979). **On law and ideology**. Basingstoke: Macmillan, p. 65. Cited in Hall, S. (1996). Introduction: Who Needs 'Identity'? In S. Hall and P. duGay (Eds.), **Questions of Cultural Identity**. London: Sage.

Hobson, D. (1982). ***Crossroads:*** **The drama of a soap opera**. London: Methuen.

Hoge, D., Johnson, B., & Luidens, D. (1994). **Vanishing boundaries : The religion of mainline Protestant baby boomers**. Louisville, Ky.: Westminster/John Knox Press.

hooks, b. (1990). **Yearning: Race, gender and cultural politics**. Boston: South End Press.

Hoover, S. (forthcoming). The cross at Willow Creek. In J. Mahan and B. Forbes (Eds.), **Religion and Popular Culture**. Manuscript submitted for publication.

Hoover, S. (1998). **Religion and the Press**. London: Sage.

Hoover, S. (1996). Media and the construction of the religious public sphere. In Hoover, S. and Lundby, K. **Rethinking media, religion and culture**. Thousand Oaks: Sage.

Hoover, S. (1995). Religion in Public Discourse: The role of the media. Mimeographed final report to the Lilly Endowment Inc., Center for Mass Media Research, University of Colorado, Boulder.

Hoover, S. (1988). **Mass Media Religion**. Thousand Oaks: Sage.

Hoover, S. and Venturelli, S. (1996). The category of the religious: The blindspot of contemporary media theory? **Critical studies in mass communication** 13, 251-265.

Hunter, J.D. (1991). **Culture wars: The struggle to define America**.

New York: Basic books.

Huyssens, A. (1984). Mapping the post-modern. **New German Critique**, 33, 5-52.

Iannacone, L. (1991). The consequences of religious market structure. **Rationality and Society** 3, 156-177.

Jacobs, M. (1992). **Sigmund Freud**. London: Sage.

Jameson, F. (1991). **Postmodernism, or the logic of late capitalism**. London: Verso.

Jenkins, H. (1988). Star Trek rerun, reread, rewritten. **Critical Studies in Mass Communication** 5(2), 85-107.

Jensen, J. (1990). **Redeeming modernity: Contradictions in media criticism**. Newbury Park, CA: Sage.

Jensen, K.B. (1990). The politics of polysemy: Television news, everyday consciousness, and political action. **Media, Culture, and Society** 12, 57-77.

Jensen, K.B. (1986). **Making sense of the news: Towards a theory and an empirical model of reception for the study of mass communications**. Aarhus: Aarhus University Press.

Jensen, K.B. (1995). **The social semiotics of mass communication**. London: Sage.

Jensen, K.B. (1996). Respondent comments, working roundtable on theory and methodology in current reception research: Studies of meaning and ontology. Presented at the joint session of the Network of qualitative audience researchers and the media, myth and ritual interest group of the International Association of Mass Communication Researchers, Sydney, August.

Jhally, S. and Lewis, J. (1992). **Enlightened racism**. Boulder: Westview press.

Jindra, M. (1994). Star Trek Fandom as a religious phenomenon. **Sociology of Religion** 55(1), 27-51.

Johnson, O. (Ed). (1993). **Information please almanac, 1994**. Boston, MA: Houghton, Mifflin Co. Cited in Chadwick, B. and Heaton, T. (Eds.)

(1996). **Statistical handbook on adolescents in America.** Phoenix: Oryx Press.

Katz, E. (1959). Mass communication research and the study of culture. **Studies in public communication** 2, 1-6. Cited in McQuail, D. (1984). With the benefit of hindsight: Reflections on uses and gratifications research. **Critical studies in mass communication** 1 (2, June), 177-193.

Kellner, D. (1992). Popular culture and the construction of postmodern identities. In Lash, S. and Friedman, J. (Eds.), **Modernity and Identity.** Cambridge, MA: Blackwell, 141-177.

Kohlberg, L. (1958). **The development of modes of moral thinking and choice in the years 10 to 16.** Unpublished doctoral dissertation, University of Chicago.

Krueger, R. (1994). **Focus groups: A practical guide for applied research,** 2nd ed. Thousand Oaks: Sage.

Kuhn, A. (1982). **Women's pictures: Feminism and cinema.** London, Boston: Routledge and Kegan Paul.

Lasch, C. (1978). **The culture of narcissism: American life in an age of diminishing expectations.** New York: Norton.

Lash, S. and Friedman, J. (1992). Introduction: Subjectivity and modernity's 'Other.' In S. Lash and J. Friedman (Eds.) **Modernity and Identity** (pp. 1-30). Oxford: Blackwell.

Lefebvre, S. (1997). Generations and religion in medias and popular culture. Paper presented to the eighth international conference on theology and communication, Gregorian University, Rome (September).

Lemert, C. (Ed.) (1993). **Social theory: The multicultural and classic readings.** Boulder: Westview Press.

Liebes, T. & Katz, E. (1990). **The export of meaning.** Oxford: Oxford University Press.

Lincoln, Y. & Guba, E. (1985). **Naturalistic Inquiry.** Beverly Hills, CA: Sage.

Linderman, A. (1997). Quantitative vs. qualitative methods in the study of religious television in Sweden. Paper presented to the Society for the Scientific Study of Religion, San Diego, CA.

Linderman, A. (1995). The reception of religious television: Social semieology applied to an empirical case study. Uppsala: Acta Univ. Uppsala. Psychologia et Sociologia Religionum 12.

Lindlof, T. (1995). **Qualitative communication research methods.** Thousand Oaks: Sage.

Lindlof, T. and Grubb-Swetnam, A. (1996). Seeking a path of greatest resistance: The self becoming method. In D. Grodin and T. Lindlof (Eds.), **Constructing the Self in a Mediated World.** Thousand Oaks, CA: Sage.

Lindlof, T. and Meyer, T. (1987). Mediated communication as ways of seeing, acting, and constructing culture: The tools and foundations of qualitative research. In Lindlof, T.(Ed.), **Natural audiences: Qualitative research of media uses and effects** (pp. 139-160). Norwood, NJ: Ablex Publishing.

Lippy, C.H. (1994). **Being religious, American style: A history of popular religiosity in the United States.** Westport, CN: Greenwood Press.

Locke, J. (1968). Some thoughts concerning education. In J.L. Axtell (Ed.), **Educational Writings.** London: Cambridge University Press.

Lopez, M. (1992). The post-work society in Puerto Rico. Lecture presented at Department of Puerto Rican and Hispanic Carribbean Studies, Rutgers University. Cited in Katz, C. (1994). Playing the field: Questions of fieldwork in geography. **The Professional Geographer** 46(1), 67-72.

Luckmann, T. (1991). Religion old and new. In P. Bourdieu and J. Coleman (Eds)., **Social theory in a changing society** (pp. 167-182). Boulder, CO: Westview.

Lull, J. (1988), **World families watch television.** Newbury Park, CA: Sage.

Lull, J. (1980). The social uses of television. **Human Communication Research** 6(3), 197-209.

Lull, J. (1995). **Media, communication, culture: A global approach.** New York: Columbia University Press.

Lunt, P. and Livingstone, S. (1996). Rethinking the focus group in media and communications research. **Journal of Communication** 46 (2), 79-98.

Lyotard, J. (1984). **The postmodern condition: A report on knowledge.** Manchester: Manchester University Press.

Malinowski, B. (1922). **Argonauts of the western pacific.** Prospect Heights, IL: Waveland Press.

Marcia, J.E. & Friedman, M.J. (1970). Ego identity status in college women. **Journal of Personality** 38, 249-263.

Martin-Barbero, J. (1993). **Communication, culture and hegemony: From the media to mediations.** (E.Fox & R.A. White, Transl.) Mexico City: Gustavo Gilli.

Marty, M. (1996. January). **Giving shape to Proteus.** Keynote address presented at the Public Conference on Media, Religion and Culture, Boulder, CO.

Marx, K. & Engels, F. (1974/1846), **The German Ideology** (student edition). Arthur, C.J. (ed. and transl.). London: Lawrence and Wishart.

Marx, K. and Engels, F. (1977). **Selected Letters.** Peking: Foreign Languages Press.

McDannell, C. (1995). **Material Christianity.** New Haven: Yale University Press.

McGuigan, J. (1992). **Cultural populism.** London: Routledge.

McRobbie, A. (1991). **Feminism and youth culture: From 'Jackie' to 'Just Seventeen.'** Boston: Unwin Hyman.

McNay, L. (1994). **Foucault: A critical introduction.** Cambridge: Polity Press.

McRobbie, A. & Gerber, J. (1978). Girls and subcultures: An exploration. In S. Hall & T. Jefferson (Eds.), **Resistance through Rituals: Youth Subcultures in Post-War Britain** (pp. 209-222). London: Hutchinson.

Mead, G.H. (1910/1964). What social objects must psychology presuppose? **Journal of philosophy, psychology, and scientific methods,** 7, 174-180. Cited in Denzin, N. (1992). **Symbolic interactionism and cultural studies.** Oxford: Blackwell.

Mead, M. (1928). **Coming of age in Samoa.** An Arbor: Morrow.

Merton, R.K., Fiske, M., and Curtis, A. (1956). **The focused interview.** New York: Free press. Cited in Lunt, P. and Livingstone, S. (1996). Rethinking the focus group in media and communications research. **Journal of Communication** 46 (2), 79-98.

Metz, C. (1982). **The imaginary signifier: Psychoanalysis and the cinema.** Bloomington: Indiana University Press.

Meyrowitz, J. (1985). **No sense of place: The impact of electronic media on social behavior.** New York: Oxford University Press.

Mills, C.W. (1963). **Power, politics and people: The collected essays of C. Wright Mills.** New York: Ballantine.

Mitchell, T. (1996). **Popular music and local identity: Rock, pop and rap in Europe & Oceania.** Leicester University Press.

Modell, J. (1989). **Into one's own: From youth to adulthood in the United States 1920-1975.** Berkeley: University of California Press.

Moore, K.A. (1992). **National commission on children: 1990 surveys of parents and children.** (Data set 19, H.M. Daley, E.C. Peterson, E.L. Lang, and J.J. Card, Archivists) [machine-readable data file and documentation]. Washington, DC: Child trends, Inc. (producer). Los Altos, CA: Sociometrics corporation, American family data archive (producer and distributor). In B.A. Chadwick & T.B. Heaton (Eds.)., **Statistical handbook on adolescents in America.** Phoenix: Oryx Press.

Moores, S. (1995). **Interpreting audiences: The ethnography of media consumption.** London: Sage.

Morgan, M., Alexander, A., Shanahan, J. & Harris, C.. (1990). Adolescents, VCRs, and the family environment. **Communication Research 17** (1), 83-106.

Morley, D. (1980). **The *Nationwide* audience: Structure and decoding.** London: BFI.

Morley, D. (1986). **Family television: Cultural power and domestic leisure.** London: Comedia.

Morley, D. and Robins, K. (1995). **Spaces of identity.** London: Routledge.

Morris, M. (1990). Banality in cultural studies. In P. Mellencamp (Ed.), **Logics of television.** Bloomington: Indiana University Press.

Morrow, R. A. (1994). **Critical theory and methodology.** Thousand Oaks, CA: Sage.

Mukerji, C. and Schudson, M. (1991). Introduction: Rethinking popular culture. In **Rethinking popular culture: Contemporary perspectives in cultural studies** (pp. 1-61). Berkeley: University of California Press.

Murphy, P. (1996). **Television and popular culture in central Mexico: An audience ethnography.** Unpublished doctoral dissertation, Ohio University, Athens.

Mulvey, L. (1975). Visual pleasure and narrative cinema. **Screen,** 16(3), 6-18.

Nagar, R. (1997). Exploring Methodological Borderlands through Oral Narratives. In J.P. Jones, H. Nast, and S. Roberts (Eds.), **Thresholds in Feminist Geography: Difference, Methodology, Representation.** London: Rowman and Littlefield.

Naficy, H. (1993). **The making of exile cultures: Iranian television in Los Angeles.** Minneapolis: University of Minnesota Press.

Nagar, R. (forthcoming). Contesting social relations in communal places: Identity politics among Asian communities in Dar es Salaam. In Archer, R. and Jacobs, J. (Eds.), **Cities of Difference.** New York: Guilford Press.

Nagar, R. (1995). **Making and breaking boundaries: Identity politics among south Asians in postcolonial Dar Es Salaam.** Unpublished dissertation, University of Minnesota.

Natrajan, B. & Parameswaran, R. (1997). **Journal of communication inquiry** 21(1), 27-59.

Newcomb, H. & Hirsh, P. (1994). Television as a cultural foum. In Newcomb, H. (Ed.), **Television: The critical view,** 5th ed. (pp. 503-515). Oxford University Press.

Ortner, S. (1993). Ethnography among the Newark: The class of '58 of Weequahic high school. **Michigan Quarterly Review** 32 (3), 411-429.

Ortner, S. (1984). Theory in anthropology since the sixties. **Comparative Studies in Society and History** 26, 126-166.

O'Sullivan, T., Hartley, J., Saunders, D. & Fiske, J. (1983). **Key concepts in communication.** London: Methuen.

Park, R. (1925/1967) Community organization and juvenile delinquency. In Park, R. and Burgess, E.W. (Eds.) **The City.** Chicago: University of Chicago Press.

Parks, S. (1986). **The critical years: The young adult search for a faith to live by.** San Francisco: Harper and Row.

Patton, M. (1990). **Qualitative evaluation and research methods,** 2nd ed. Newbury Park, CA: Sage.

Plato. (1987). From **The Republic.** Trans. Desmond Lee, 2nd ed. (revised). London: Penguin.

Peck, J. (1994). Talk about racism: Framing a popular discourse of race on *Oprah Winfrey.* **Cultural Critique** 27, 89-126.

Pearl, D., Bouthilet, L., & Lazar, J. (Eds.). (1982). **Television and behavior: Ten years of scientific progress and implications for the eighties** (Vol. 1, summary report). Rockville, MD: National Institute of Mental Health.

Personal Narratives Group (Eds.). (1989). **Interpreting women's lives: Feminist theory and personal narratives.** Bloomington: Indiana University Press.

Piaget, J. (1985). **The equilibration of cognitive structures: the central problem of intellectual development.** Transl. T. Brown & K. J. Thampy. Chicago: University of Chicago Press.

Piaget, J. (1965). **The moral judgment of the child.** Transl. M. Gabain. New York: Free Press.

Piaget, J. & Inhelder, B. (1969). **The psychology of the child.** Transl. H. Weaver. New York: Basic Books.

Pipher, M. (1994). **Reviving Ophelia: Saving the selves of adolescent girls.** New York: Ballantine books.

Postman, N. (1985). **Amusing ourselves to death.** New York: Penguin.

Powell, J. (1997). The TV show that works miracles: The top-rated series. **Good Housekeeping** 224(4), 108-110.

Press, A. (1991). **Women watching television: Gender, class and generation in the American television experience.** Philadelphia: University of Pennsylvania Press.

Pfister, J. and Schnog, N. (Eds.) (1997). **Inventing the psychological: Toward a cultural history of emotional life in America.** New Haven: Yale University Press.

Radway, J. (1987). **Reading the romance: Women, patriarchy and popular literature.** London: Verso.

Rakow, L. (1988). Gendered technology, gendered practice. **Critical studies in mass communication 5** (1), 57-70.

Rakow, L. & Navarro, V. (1993). Remote mothering and the parallel shift: Women meet the cellular telephone. **Critical studies in mass communication 10** (2), 144-157.

Robinson, D., Buck, E., & Cuthbert, M. (1991). **Music at the margins.** London: Sage.

Roe, K. (1983). **Mass media and adolescent schooling: Conflict or co-existence?** Stockholm, Sweden: Almqvist & Wiksell International.

Roof, W.C. (1993). **A Generation of Seekers: The spiritual journeys of the baby boom generation.** San Francisco: Harper.

Roof, W.C. (1997). **Religious borderlands: Challenges for future study.** Presidential address to the annual meeting of the Society for the Scientific Study of Religion, San Diego.

Roof, W.C. (1982). America's voluntary establishment: Mainline religion in transition. In M. Douglas and S. Tipton (Eds.), **Religion and America** (pp. 130-149). Boston: Beacon.

Rosengren, K.E. (1983). Communication research: One paradigm, or four? **Journal of Communication 33,** p. 203, cited in Ang, I. (1996). **Living Room Wars: Rethinking media audiences for a postmodern world.** London: Routledge.

Rosengren, K.E. & Windahl, S. (1989). **Media Matter: TV use in**

childhood and adolescence. Norwood, NJ: Ablex Publishing.

Rowland, W. (1983). **The politics of TV violence: Policy uses of communication research.** Beverly Hills: Sage.

Sampson, E. (1989). The deconstruction of the self. In J. Shotter and K. Gergen (Eds.), **Texts of identity** (pp.1-19). London: Sage.

Saussure, F. (1917/1976). **Course in general linguistics.** (W. Baskin, Trans). New York: McGraw Hill.

Schlegel, A. & Barry, H. III. (1991). **Adolescence: An anthropological inquiry.** New York: Free Press.

Schudson, M. (1977). **Discovering the news: a social history of American newspapers.** New York: Basic books.

Schultze, Q., Anker, R., Bratt, J., Romanowski, W., Worst, J.,& Zuidervaart, L. (1991). **Dancing in the dark: Youth, popular culture and the electronic media.** Grand Rapids, MI: William B. Eerdmans.

Selley, A. (1990). Transcendentalism in *Star Trek: The Next Generation.* **Journal of American Culture** 13,31-34.

Shelton, C. (1983). **Adolescent spirituality.** Chicago: Loyola University Press.

Shohat, E. (1993). Gender and the culture of empire: Toward a feminist ethnography of cinema. In Naficy, H. and Gabriel, T. (Eds.) **Otherness and the media: The ethnography of the imagined and the imaged.** Chur, Switzerland: Harwood academic publishers.

Shohat, E. & Stam, R. (1994). **Unthinking Eurocentrism: Multiculturalism and the media.** London: Routledge.

Shotter, J. and Gergen, K. (Eds.). (1989). **Texts of identity.** London: Sage.

Silverstone, R. (1981). **The message of television.** Bedford Square, England: Henemann Educational Books.

Silverstone, R. , Hirsch, E., Morley, D. (1991). Listening to a long conversation: An ethnographic appraoch to the study of information technologies in the home. **Cultural Studies** 5(2), 204-27.

Singer , J.L. & Singer, D.G. (1981). **Television, imagination and aggression: A study of preschoolers.** Hillsdale, NJ: Lawrence Erlbaum.

Slugorski, B.R. & Ginsberg, G.P. (1989). Ego identity and explanatory speech. In J. Shotter and K. Gergen (Eds.), **Texts of identity** (pp. 36-55). London: Sage.

Somers, M. (1992). Narrativity, narrative identity and social action: Rethinking English working-class formation. **Social science history** 16(4), 591-630.

Snyder, T.D., & Hoffman, C.M. (1993). **Digest of education statistics.** Washington, DC: National center for education studies.

Stabile, C. (1995). Resistance, recuperation, and reflexivity: The limits of a paradigm. **Critical studies in mass communication** 12, 403-422.

Stacey, J. (1991). **Brave new families: Stories of domestic upheaval in late twentieth century America.** New York: Basic.

Storey, J. (1993). **An introductory guide to cultural theory and popular culture.** Athens: University of Georgia Press.

Strasburger, V. (1995). **Adolescents and the media: Medical and psychological impact.** Thousand Oaks, CA: Sage.

Toqueville, A. (1969). **Democracy in America.** Transl. G. Lawrence, J.P. Mayer (Ed.). New York: Doubleday.

Tschannen, O. (1994). Sociological controversies in perspective. **Review of religious research** 36(1), 70-86.

Tsing, A. L. (1993). **In the realm of the diamond queen.** Princeton University Press.

Tuchman, G. (1978). **Making news: A study in the construction of reality.** New York: Free Press.

Tucker, K. (1997). Salvation army. **Entertainment Weekly** 383 (2), 46-47.

Turner, G. (1992). **British cultural studies: An introduction.** London: Routledge.

Turner, V. (1974). **Dramas, Fields, and Metaphors.** Ithaca, NY: Cornell

University Press.

U.S. Bureau of the Census (1993). Statistical abstract of the United States (113 edition), Washington, D.C: U.S. Government Printing Office.

U.S. Department of Health and Human Services, National Center on Child Abuse and Neglect. (1994). **Child maltreatment 1992: Reports from the states to the National Center on Child Abuse and Neglect.** Washington, DC: U.S. Government Printing Office.

VanMaanen, J. (1988). **Tales of the field: On writing ethnography.** Chicago: University of Chicago Press.

Van Zoonen, L. (1994). **Feminist media studies.** London: Sage.

Visweswaran, K. (1994). **Fictions of feminist ethnography.** Minneapolis: University of Minnesota Press.

Walkerdine, V. (1986). Video replay: Families, films, and fantasy. In V. Burgin, J. Donald, and C. Kaplan (Eds.), **Formations of fantasy** (pp. 167-199). London: Methuen.

Warner, S. (1993). Work in progress toward a new paradigm for the sociological study of religion in the United States. **American Journal of Sociology** 98(5), 1044-1093.

Warren, M. (1992). **Communications and cultural analysis: A religious view.** Westport, CN: Bergin & Garvey.

Warren, M. (1987). **Youth, gospel, liberation.** San Francisco: Harper and Row.

Warren, M. (1982). **Youth and the future of the church: Ministry with youth and young adults.** New York: Seabury Press.

Wartella, E. (1980). Children and television: The development of the child's understanding of the medium. In G.C.Wilhoit and H. deBock (Eds.), **Mass communication review yearbook.** Beverly Hills, CA: Sage.

White, R. (1994). Audience 'interpretation' of media: Emerging perspectives. **Communication research trends** 14 (3), 1-40.

Whitlock, K. (1989). **Bridges of respect: Creating support for lesbian and gay youth,** 2nd ed. Philadelphia: American Friends Service Committee.

Wilcox, M. (1979). **Developmental journey.** Nashville: Abingdon.

Williams, P. (1980). **Popular religion in America.** Englewood Cliffs, NJ: Prentice-Hall.

Williams, P. and Chrisman, L. (1994). Colonial discourse and post-colonial theory: An introduction. In P. Williams and L. Chrisman (Eds.), **Colonial discourse and post-colonial theory: A reader** (pp. 1-20). New York: Columbia University Press.

Williams, R. (1977). **Marxism and literature.** London: Oxford University Press.

Winn, Marie. (1985). **The plug-in drug,** rev. ed. New York: Viking Press.

West, C. (1993). **Race matters.** Boston: Beacon Press.

Willis, P. (1977). **Learning to labor: How working class kids get working class jobs.** London: Gower.

Wright, S.A. (Ed.). (1997). Mass media and unconventional religion (Special issue). **Review of religious research** 39(2).

Wuthnow, R. (1995). **Learning to care: Elementary kindness in an age of indifference.** New York/Oxford: Oxford University Press.

Wuthnow, R. (1988). **The restructuring of American religion.** Princeton University Press.

Wuthnow, R. (1992). **Rediscovering the sacred: Perspectives on religion in contemporary society.** Grand Rapids: Eerdmans.

Yamane, D. (1997). Secularization on trial: In defense of a neosecularization paradigm. **Journal for the scientific study of religion** 36 (1), 109-122.

Yin, R. (1994). **Case study research: Design and methods,** 2nd ed. Thousand Oaks, CA.

Zillmann, D. (1971). Excitation Transfer in communication-mediated aggresive behavior. **Journal of experimental and social psychology** 7, 419-434

Appendix: List of Research Participants

Teens involved in case study research who served as peer-led discussion group leaders:

Elizabeth Farley
14 - 16 years old
Caucasian
working class
active member, Lutheran church (ELCA)

*****

Michael Pearson
15 - 16 years old
African American
working class
attends Church of Christ

*****

Hasan Ahmed
14 - 15 years old
Arab African/Caucasian
middle class
active member, Colorado Muslim Society

Other teens involved in case study research:

Amber Dearborn
16 years old
Native American/Caucasian
working class
attends Baptist church

*****

Jake Pickerington
17 years old
Caucasian
upper middle class
no religious background or affiliation

Parents of teens who organized and participated in discussion groups:

Judy and Tom Ortiz, organizers
Mexican American
middle class
active in Roman Catholic church
teen daughter, teen son

Connie
Mexican American
Judy's older sister
lower middle class
active in Roman Catholic church
single parent, one teenage son; one who died a few years ago as teen

Cheryl and Taylor
she - Mexican American; he - African American
Judy's younger sister
working class
occasional attenders, Roman Catholic church
both single parents of teens

Rob and Diane
Mexican American
working class
active in Roman Catholic church
separated; raising teen daughter together

Cathy Swenson, organizer
Caucasian
lower middle class
active in nondenominational church
single parent, two teen daughters (only one at home)

Jean
Caucasian
working class
occasional attender, Lutheran (ELCA)
single parent, teen daughter and teen son

Sandy
Caucasian
lower middle class
active in Roman Catholic church
single parent, teen daughter

Joe
Native American
Working class
attends Bapist church
single parent, teen daughter and teen son

Donna
Caucasian
middle class
attends Roman Catholic/United Methodist churches
single parent, teen son and teen daughter

Dick
Caucasian
lower middle class
no religious affiliation
single parent, two teen sons, teen daughter

Justine
lower middle class
no religious affiliation
married, teen daughter

Lester and Ellen Gray, organizers
Caucasian
upper middle class
Presbyterian pastor (PCUSA) and active member
two teen daughters

Jack and Tanya
Caucasian
upper middle class
Presbyterian
teen daughter and teen son

Don and Julie
Caucasian
upper middle class
Presbyterian
teen son and teen daughter

Melanie
Caucasian
upper middle class
Presbyterian
married, teen daughter and teen son

Members of Teen Peer-Led Discussion Groups:

Elizabeth's Group:

Allyson
15 years old
Caucasian
middle class
no religious affiliation
lives with mother, father, two brothers, sister, cousin

Vickie
15 years old
Caucasian
middle class
occasionally attends Roman Catholic church
lives with mother, father, two sisters

Betsi
14 years old
Caucasian
middle class
occasionally attends "various churches"
lives with mother, father, two brothers

Lisa
15 years old
Caucasian
middle class
active in Presbyterian church (PCUSA)
lives with mother, father, brother

Tara
16 years old
Caucasian
middle class
no religious affiliation
lives with mother, father, brother

Michael's Group:

Daryl
14 years old
African American
working class
active in Jehovah's Witnesses
Dillon's twin brother
lives with mother, father, brother

Dillon
14 years old
African American
working class
active in Jehovah's Witnesses
Daryl's twin brother
lives with mother, father, brother

Randall
14 years old
African American
disadvantaged
no religious affiliation
lives with mother

Michael
14 years old
African American
disadvantaged
no religious affiliation
lives with mother, father, brother

Burt
14 years old
African American
disadvantaged
no religious affiliation
lives with grandmother, three uncles, two brothers

Hasan's Group:

Mitch
15 years old
Korean/Philippino
lower middle class
attends Roman Catholic church
lives with mother, father

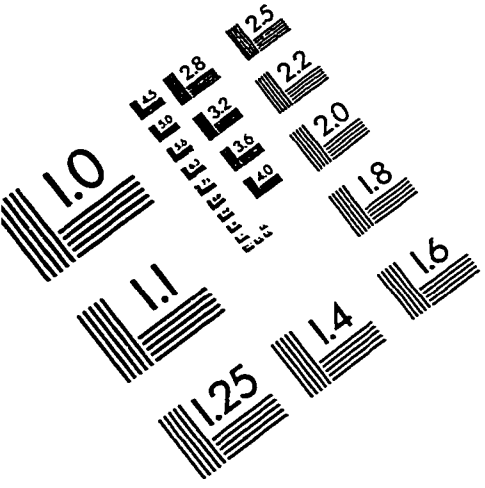
Bailey
15 years old
African American
lower middle class
attends nondenominational church
lives with mother

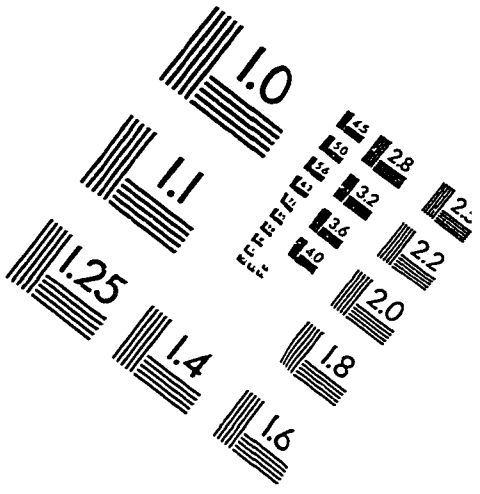
Chuck
15 years old
African American/Caucasian
lower middle class
active in Nazarene church
lives with mother, father, sister

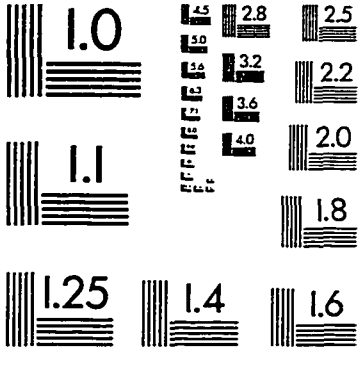
Kenny
15 years old
African American
working class
attends Roman Catholic church
lives with grandmother, aunt, uncle

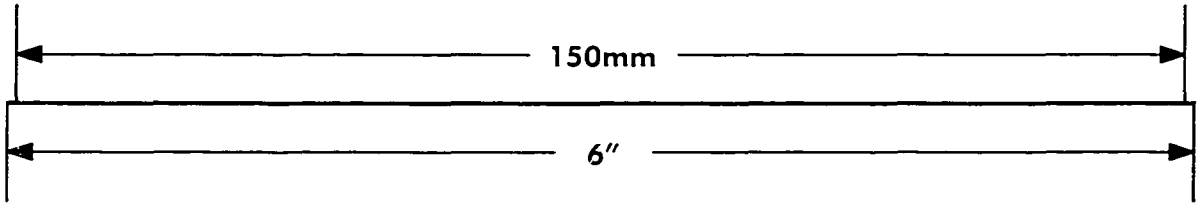
Mike
15 years old
African American
middle class
no religious affiliation
lives with mother

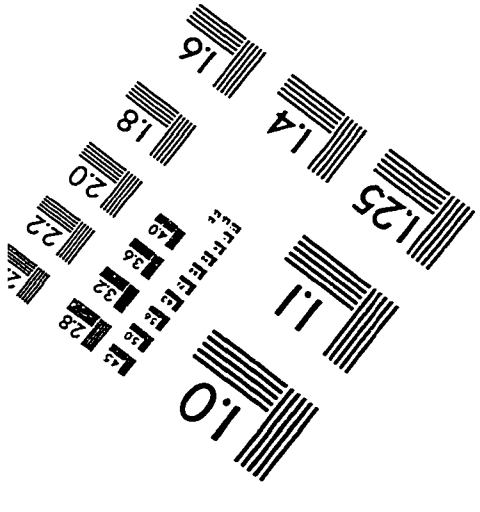
IMAGE EVALUATION TEST TARGET (QA-3)

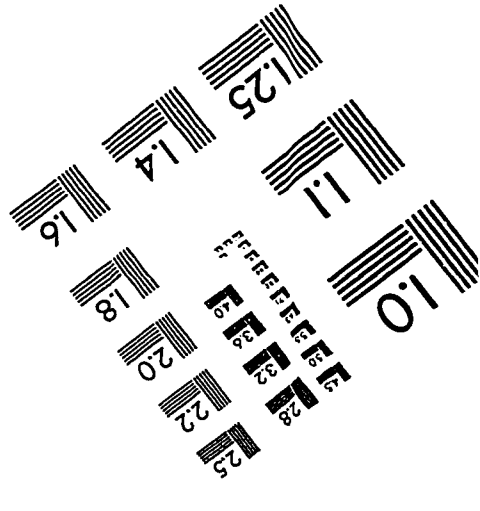












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

LONG RANGE ANISOTROPIC INTERACTIONS IN RYDBERG ATOMS AND MOLECULES

by

WILLIAM CLARK

B. S., University of Nevada, 1992

M. S., University of Colorado, 1994

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

1998

This thesis for the Doctor of Philosophy degree by

William Clark

has been approved for the

Department of

Physics

by

Chris H. Greene

Robert Parson

Date March 6, 1998

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Clark, William (Ph. D., Physics)

Long Range Anisotropic Interactions in Rydberg Atoms and Molecules

Thesis directed by Professor Chris H. Greene

The study of subtle interactions in atomic and molecular systems has stirred scientific interest since the dawn of quantum mechanics. Even today, research into perturbative long range interactions continues to push into new territory, largely driven by the experimental capabilities of Rydberg state spectroscopy. However, theoretical investigations have also made significant contributions, suggesting that the motion of a charged particle in the field of an anisotropic core is far more complicated than previously thought.

In this work we present a new theoretical formulation of Rydberg atoms and molecules that confirms the existence of previously unknown interactions, including an unusual pseudo-vector interaction with an "orbit-orbit" operator structure. While static long-range multipole interactions have been studied for years, the presence of dynamic terms that involve both position and momentum operators has been demonstrated only recently. In contrast to the ordinary quadrupole or induced-dipole interaction terms, the existence of this vector interaction hinges on the motion of the distant charge as it roams far beyond the confines of the core. Physically, this interaction reflects an attempt of the Rydberg electron to "drag" the core polarization vector with it. This drag is hindered by the internal moment of inertia of the ion, and by the moment of inertia of the distant electron about the center-of-mass .

The theoretical work described below, for high-ℓ Rydberg states of Ne, in combination with experimental work performed elsewhere, has confirmed for the first time the existence of this vector interaction. Our analysis shows that details of the spectrum, at sub-MHz resolution, depend strongly on the tensorial structure of this unusual interaction. Our long range multichannel formulation has now been applied successfully to describe autoionization resonances in doubly excited Rydberg states of Mg, and, most recently, to describe

high-ℓ Rydberg states of H_2 and D_2. This work also presents the first complete adiabatic calculations of the ground ($\nu^+ = 0, N^+ = 0$) and first rotationally excited ($\nu^+ = 0, N^+ = 1$) state polarizabilities and hyperpolarizabilities of H_2^+ and D_2^+. An analysis of the energy shifts generated by the vector interaction and the relativistic retardation (or Casimir) interaction demonstrates that they cause clearly different patterns of energy level splittings that should be experimentally observable. This theoretical work, in combination with still more sensitive experiments that are currently underway, should reveal the quantitative nature of both of these subtle interactions in a Rydberg molecule.

DEDICATION

I dedicate this work to my niece Anastasia Clark King who has been a great source of love and inspiration, and who reminds me every day that life is an adventure that is to be enjoyed. Thanks Pooh!

ACKNOWLEDGEMENTS

I first want to thank my advisor Professor Chris H. Greene for his support over the past few years, support that has allowed me to mature as a scientist and that has allowed me to interact with some of the best people in this field. The work presented here originated from discussions with Chris regarding the work of Stephen R. Lundeen and Bernard Zygelman during the previous three years, and has developed into a complete theoretical description of Rydberg atoms and molecules that provides both spectroscopic accuracy and physical insight.

Secondly, I want thank Professor Bernard Zygelman for providing me with encouragement and enthusiasm throughout my undergraduate and graduate years. I thank Bernard for having genuine interest in me and my future, and for taking the time to make a difference. I also want to thank Professor Thomas Degrand, who like Bernard, has encouraged me and impressed me with his exceptional knowledge, intelligence, wit, and advise.

Thirdly, I want to thank the numerous people I have had the privilege to interact with at JILA. I thank Robert P. Wood, Kurt W. Meyer, Gregory Miecznik, Jeff Stephens, Brett D. Esry, James P. Burke, Stephanie Staley, Andrew Hooker, John L. Bohn, Brian Eleson Granger, and Hugo W. Van der Hart whom made up the Greene Research Group during my graduate years. I also thank, Professor Robert Parson and my other committee members for reading this manuscript and provided many useful suggestions.

Finally, I want to thank the National Science Foundation for supporting my research interests, for allowing me to publish my work in peer reviewed, international journals, and for making it possible to attend and present my research at international conferences.

CONTENTS

CHAPTER

1 INTRODUCTION 1

2 ADIABATIC DESCRIPTION OF RYDBERG SYSTEMS 8

2.1 Adiabatic Representation 8

2.2 Adiabatic Close-Coupling Equations 10

2.3 Adiabatic Potentials and Post-Adiabatic Corrections 12

2.4 Perturbative Adiabatic Diagonalization for Nondegenerate Channels . . 14

2.5 Degenerate Channel Contributions 17

2.6 Advantages and Disadvantages 18

3 DIABATIC FORMULATION OF THE RYDBERG CHANNEL INTERACTIONS 19

3.1 Diabatic Representation 19

3.2 Diabatic Close-Coupling Equations 20

3.3 Generalized Eigenvalue Problem 23

3.4 Autoionization Rates 26

3.5 Advantages 27

4 RECOUPLING AND PARAMETRIZATION 29

4.1 Motivation 29

4.2 Recoupling of Spherical Tensor Operators 31

4.2.1 Recoupling of First-Order Terms 32

4.2.2 Recoupling of Second-Order Terms 33

4.2.3 Unit Tensor Notation 34

4.3 Parametrization of Interactions 35

4.4 Tensor Analysis of Adiabatic Potential 36
4.4.1 Nondegenerate Channel Contributions 37
4.4.2 Degenerate Channel Contributions 41
4.5 Tensor Analysis of Diabatic Hamiltonian 43
4.5.1 High ℓ States of Low Z Atoms 44
4.6 Qualitative Interpretation of the Vector Interaction 46
4.6.1 Analytic Recoupling Analysis 46
4.6.2 A Two-Electron Example Treated Classically 48
4.6.3 Two-Electron Example Treated Using Quantum Mechanics . . . 49
4.7 Non-Abelian Gauge Formulation 53
5 RYDBERG STATES OF NEON 54
5.1 Application to Neon 54
5.2 First Calculation of the Vector Hyperpolarizability β_v 60
6 AUTOIONIZING RYDBERG STATES OF MAGNESIUM 67
6.1 Modification for Low ℓ Rydberg States 67
6.2 Radial Dependence of Core Parameters 69
6.3 Rydberg Levels and Rates: Mg $3pnf$ 72
6.4 Adiabatic Torquing of Orbital Planes 80
7 RYDBERG STATES OF THE HYDROGEN DIATOMIC MOLECULE . . . 87
7.1 Molecular Hamiltonian in Jacobi Coordinates 87
7.1.1 Transformation 88
7.1.2 Kinetic and Potential Operators 89
7.1.3 Molecular Ion and Rydberg Hamiltonians 89
7.1.4 Spherical Expansion of the Rydberg Hamiltonian 90
7.2 Close-Coupling Representation 92
7.3 Levels of Approximation 94

7.3.1 Born-Oppenheimer Approximation 95

7.3.2 Adiabatic Nuclei Approximation 95

7.4 Core States of H_2^+ and D_2^+ 96

7.4.1 Calculation of Electronic States 96

7.4.2 Calculation of Rovibrational States 98

7.4.3 Core Parameters and the Vector Hyperpolarizability 98

7.5 Rydberg States of H_2 and D_2 104

7.5.1 Ryberg States of ($\nu^+=0, N^+=0$) H_2 and D_2 104

7.5.2 Analysis of the Vector Interaction and Relativistic Retardation in the $n = 10$ Rydberg States of ($\nu^+ = 0, N^+ = 1$) H_2 106

8 SUMMARY AND REMARKS 115

BIBLIOGRAPHY 118

APPENDIX

A PARAMETERS FOR HIGH-ℓ RYDBERG STATES OF LOW-Z ATOMS . 124

B CHANNEL PARAMETERS FOR RYDBERG STATES OF H_2 AND D_2 . . 128

C STURMIAN BASIS 134

D PROLATE SPHEROIDAL COORDINATES 136

E FINITE ELEMENT METHOD 138

FIGURES

FIGURE

3.1 Radial potential curves. 21

4.1 Classical demonstration of vector interaction using orbits 51

4.2 Perturbation in energies of orbits 52

5.1 Adiabatic potential curves for $K^\pi = \frac{9}{2}^-$ Ne. 61

5.2 Adiabatic potential curves for $K^\pi = \frac{11}{2}^-$ Ne 62

6.1 Channel-independent radial parameters for the Mg^+ $3p$ ion. 73

6.2 Mg $3p_{\frac{1}{2}} nf_{\frac{7}{2}} (J = 4)$ autoionization rates versus n. 77

6.3 Mg $3p_{\frac{3}{2}} nf_{\frac{7}{2}} (J = 4)$ autoionization rates versus n. 78

6.4 Mg $3p_{\frac{3}{2}} nf_{\frac{9}{2}} (J = 4)$ autoionization rates versus n. 79

6.5 Adiabatic potential curves for $K^\pi = \frac{7}{2}^-$ Mg. 81

6.6 Adiabatic potential curves for $K^\pi = \frac{9}{2}^-$ Mg. 82

6.7 Orientation of orbital planes for $K^\pi = \frac{7}{2}^-$ Rydberg Mg 84

6.8 Orientation of orbital planes for $K^\pi = \frac{9}{2}^-$ Rydberg Mg 85

7.1 Comparison of N-dependent tensor energy patterns. 110

7.2 Vector and Casimir energy shifts in H_2. 112

7.3 Vector and Casimir energy shifts in H_2 versus N. 113

7.4 Vector and Casimir energy shifts in H_2 versus L. 114

E.1 Polynomial basis set used in finite element method. 140

TABLES

TABLE

5.1 Experimental and theoretical parameters for Ne^+. 57

5.2 Experimental and theoretical $n = 10$ Rydberg intervals in Ne. 58

5.3 Theoretical and experimental energies of of some Ne^{2+} states. 65

6.1 $e - Mg^{++}$ model potential parameters. 69

6.2 Mg^+ $3p$ ionic core parameters. 71

6.3 Mg $3p_{\frac{1}{2}}nf_{\frac{7}{2}}J = 4$ levels, rates, and quantum defects. 75

6.4 Mg $3p_{\frac{3}{2}}nf_{\frac{7}{2}}J = 4$ levels, rates, and quantum defects. 75

6.5 Mg $3p_{\frac{3}{2}}nf_{\frac{9}{2}}J = 4$ levels, rates, and quantum defects. 76

7.1 Ground states polarizabilities and hyperpolarizabilities of H_2^+ and D_2^+. . . 101

7.2 Multipole moments and polarizabilities for $(v^+ = 0, N^+ = 1)$ H_2^+ 105

7.3 Comparison of energy shifts for various powers of $\frac{1}{r}$ 107

7.4 $n = 10$, $(v^+ = 0, N^+ = 0)$ Rydberg intervals for H_2 and D_2. 108

CHAPTER 1

INTRODUCTION

The subtle influences of tiny interactions have frequently served as probes of new physics and as a reminder of the complexity of nature. The discrepancy between the observed and calculated perihelion of the planet Mercury, a mere difference of 43″ of arc per century, was the outstanding problem in Newtonian physics until the equations of planetary motion were modified by the general theory of relativity. The existence of the Lamb shift in the $2S_{\frac{1}{2}}$ and $2P_{\frac{1}{2}}$ states of hydrogen ($\approx$ 1057MHz) [1, 2] was found to originate from the interaction between the bound electron and vacuum fluctuations, and lead to the first systematic removal of logarithmic divergences from early relativistic theories and to the creation of renormalization techniques in modern quantum field theory [3, 4, 5, 6, 7]. Even the spin of an electron, while in some way a subtle atomic property, was originally discovered through its effects in spectroscopy, and has had far reaching implications including the number of electrons that can occupy a give quantum state (Pauli's exclusion principle) and of course the basic chemical properties of atoms and molecules.

Since the first measurement of the Lamb shift [1, 2], a half-century ago, the ability of experiments to probe the weakest forces in nature has grown enormously. Today the electroweak interaction and the nuclear anapole moment can be measured in table top atomic physics experiments [8], thereby broadening the traditional scope of atomic physics and complementing studies in high energy and nuclear physics. The field of Rydberg state spectroscopy [9, 10, 11, 12, 13, 14, 15, 16, 17] has likewise grown tremendously, providing a revealing probe of the subtle interactions between atomic and molecular ionic cores and weakly bound Rydberg electrons. High resolution microwave spectroscopy can now resolve

Rydberg intervals with a precision of 1 kHz [9, 10, 11], yielding precise, indirect measurements of ionic core multipole moments and polarizabilities. And, as a consequence of these improvements in resolution, experiments are close to measuring, for the first time, the slight energy shifts of Rydberg states due to relativistic retardation, or "Casimir" forces [18, 19].

In recent years, new physics has been discovered in the long range interactions of anisotropic Rydberg atoms and molecules. The work presented here not only chronicles this new discovery, but essentially represents much of its history. To put this discovery in perspective let us begin by reviewing some simple features of high-ℓ Rydberg motion.

The unusual spatial nature of nonpenetrating Rydberg states provides a considerable degree of theoretical simplification. The strong centrifugal repulsion, associated with high-ℓ motion, forces the Rydberg electron to roam far beyond the confines of the core. In this distant region the electron primarily experiences a simple screened Coulomb attraction toward the ion. However, it is the structure of the ionic core that makes Rydberg physics interesting and surprisingly complicated.

In order to elucidate the importance of core structure, consider the simple case of a distant Rydberg electron attached to a spherically symmetric ground state ion, which possesses no permanent multipole moment. Apart from the long range $-\frac{1}{r}$ Coulomb attraction, the charge induces dipole moments in the ion that fall off, asymptotically, as $\frac{1}{r^2}$. The net effect of the induced dipole interaction is to further increase the interaction potential energy by $-\frac{\alpha_s}{2r^4}$, where α_s is the scalar dipole polarizability of the ionic core. As a result, the polarization interaction produces simple scalar shifts for each ℓ state, removing the hydrogenic degeneracy.

The physics of a Rydberg system changes substantially when anisotropic cores having nonzero total angular momentum are considered. The first important difference is the presence of a quadrupole Rydberg-core interaction of the form $-\frac{Q}{r^3}P_2(cos\theta)$, where θ is the relative angle between a symmetry axis of the ion and a unit vector pointing to

the Rydberg electron. Moreover, an additional tensor dipole polarization interaction of the form $-\frac{\alpha_t}{2r^4}P_2(cos\theta)$ appears that was absent in the spherically symmetric case. The tensorial dependence of these interactions produces coupling among different ionic core states and among the various Rydberg series attached to ionization thresholds of the core, which generally produces far richer and more complex spectra than are observed for systems with spherically symmetric ionic cores.

Most theoretical treatments of Rydberg systems have been restricted to interactions among charged particles and spherically symmetric closed-shell atoms or ions [20, 21, 22]. Theoretical extensions to nonpenetrating Rydberg states that involve an electron and an open-shell, anisotropic core have remained largely unexplored. A few notable exceptions in atomic and molecular physics are the perturbative polarization schemes of Schoenfeld and Sturrus [23, 13, 12, 11] and the more general nonperturbative multichannel approach of Herzberg and Jungen [24, 25, 26]. The lack of attention to anisotropic systems, and the need to better understand the long range nature of nonrelativistic physics, is highlighted by Zygelman's [27] recent prediction of a new nonrelativistic "orbit-orbit" type interaction in anisotropic Rydberg systems. In a pioneering study of long range atomic forces, Zygelman used non-Abelian gauge transformations and geometric phases to predict an effective long-range potential energy, now called the "vector" interaction, with the operator structure

$$V_{vector} = \beta_v \frac{\vec{L_c} \cdot \vec{\ell}}{r^6}. \tag{1.1}$$

Here $\vec{L_c}$ and $\vec{\ell}$ denote the orbital momentum of the ionic core and the Rydberg electron, respectively. The terminology "vector" interaction is somewhat misleading. The overall operator $\vec{L_c} \cdot \vec{\ell}$ is clearly a scalar, which conserves the total angular momentum of the Rydberg-core system. Moreover, $\vec{L_c}$ and $\vec{\ell}$ are *pseudovectors* that remain even under the spatial inversion of the core and Rydberg electronic degrees of freedom. Interestingly, this is the only nonzero term with odd tensorial structure in the effective long range potential out

to radial order $\frac{1}{r^6}$. Additional odd rank tensor terms such as the dipole or octupole terms of the Rydberg-core interaction fail to survive to this order in $\frac{1}{r}$, assuming, of course, the typical situation where ionic states possess a definite parity.

The analysis of Zygelman provided no analytic expression for the proportionality coefficient β_v, raising questions regarding not only the physical origin of this "vector" interaction, but also its existence since the coefficient could vanish. A term with this operator structure was introduced into atomic spectroscopy by Trees [28] and Racah [29], but only on semiempirical grounds and without an explicit derivation. However, they did correctly interpret this correction term as a polarization energy arising from second-order perturbative effects.

In a move to clarify the existence of this term and provide some explanation of its physical origin, we put forth a systematic derivation of this unusual interaction, using standard techniques in atomic and molecular physics. With the aid of high resolution Rydberg spectra we confirmed, for the first time, its existence in Rydberg states of Ne [30, 31, 10]. Our theory has not only established the existence of the vector interaction, but it has also spawned an extension of perturbative theories that enables us to treat effects due to coupling among Rydberg states, using a long range polarization methodology. The new multichannel Hamiltonian approach to Rydberg systems has now been successfully applied to Rydberg states of Ne [30, 31, 10], to autoionizing Rydberg states of Mg [33, 32] and, most recently, to Rydberg states of H_2 and D_2 [34, 9, 11]. Physically, the vector term is a dynamic angular interaction that reflects a coupling among the orbital motion of the Rydberg electron and the internal ionic angular momentum, mediated by the dipole portion of the electron-ion interaction. As the Rydberg electron orbits the core, it tries to drag the core polarization vector with it. This drag is hindered by the internal moment of inertia of the ion, and the moment of inertia of the distant Rydberg electron.

The theoretical description of Rydberg systems developed here differs from the

traditional perturbative approaches in two important ways. We abandon the use of Rayleigh-Schrödinger perturbation theory and approach the problem from a multichannel close-coupling perspective that is common in modern scattering theory. The multichannel formulation makes it possible to derive effective long range potentials without fully solving the complete atomic or molecular Rydberg problem. This differs from the Rayleigh-Schrödinger approaches [23, 13, 12, 11] in that we do not calculate shifts from pure hydrogenic levels, but rather systematically derive effective potentials and compute Rydberg states by diagonalizing an effective Hamiltonian. This last step provides a second important difference, in that coupling among different Rydberg states can be treated almost exactly. While it is true that the long range interaction between the Rydberg electron and the ionic core can be expanded in reciprocal powers of the Rydberg radial coordinate, Rydberg states attached to different ionization thresholds of the ionic core can still be strongly coupled to one another. Moreover, the fact that there can be near degeneracies among these strongly coupled Rydberg states invalidates the traditional nondegenerate perturbative approaches. This type of coupling is easily treated, however, with the multichannel formulation presented here.

We begin in Chapter 2 with a systematic derivation of the long range adiabatic potential between a Rydberg electron and an arbitrary anisotropic core. This derivation serves to demonstrate how a long range potential can be derived using standard techniques in atomic and molecular physics, and lays the groundwork for a theoretical prediction of the vector interaction, which is presented in Chapter 4. In Chapter 3 we generalize the single channel analysis of Chapter 2 and develop a multichannel diabatic approach to Rydberg systems that enables a systematic treatment of channel coupling. The proper treatment of channel coupling is essential in any modern theory that attempts to reproduce experimental observations with spectroscopic accuracy. In chapter 4 we use Wigner-Racah recoupling algebra to transform the potentials developed in Chapters 2 and 3 so that both their operator structure and their anisotropic nature can be easily identified. The tensor analysis also

enables the identification of channel dependent core properties that allow us to cast the long range potentials in a parametrized form. A systematic derivation of the vector interaction is then presented and its physical implications are discussed.

In Chapters 5 through 7 we apply our theoretical formulation to the description of Rydberg states in Ne, Mg, H_2, and D_2. The analysis of the $n = 10$, $\ell = 5, 6, 7$, and 8 Rydberg states of Ne, presented in Chapter 5, represents the first confirmation of the vector interaction within a Rydberg system. We show that by including the vector interaction in our analysis of the observed Ne Rydberg levels [10, 30, 31] a four fold reduction in the χ^2 comparison can be achieved, reproducing energy splittings with an accuracy of no worse that 0.5 MHz. Moreover, a comparison of our theoretical value for β_v and the value extracted from the experiment provides additional evidence for the existence of this subtle interaction.

Chapter 6 continues this investigation into anisotropic interactions by considering the doubly excited $3pnf$ Rydberg states of Mg where a Rydberg electron interacts with an excited Mg^+ $3p$ core [31, 32]. In this system the strong coupling among Rydberg states results in irregular behavior in the n-dependence of autoionization rates, which is accurately described by our diabatic formulation. While the coefficient of the vector interaction is thirty times larger for the Mg^+ $3p$ core than for the Ne^+ core, its effect is much less discernible owing to the large widths of the lower-ℓ autoionizing resonances that were examined in the Mg experiment.

Chapter 7 presents our most recent work on Rydberg states of the simple diatomic molecules H_2 and D_2. This work focuses primarily on Rydberg states attached to spherical ionic cores, namely the $(\nu^+ = 0, N^+ = 0)$ states of H_2^+ and D_2^+, since measurements with kHz resolution were recently performed by Jacobson [9]. These precise measurements give a stringent test of our diabatic formulation and our ability to accurately compute core multipole moments, polarizabilities, and hyperpolarizabilities. The first complete adiabatic calculations of polarizabilities and hyperpolarizabilities for the ground $(\nu^+=0, N^+=0)$ and

first rotationally excited ($\nu^+=0, N^+=1$) states of H_2^+ and D_2^+ are presented. The strength of our treatment is demonstrated by our ability to reproduce Rydberg intervals with a accuracy of better than 0.3 MHz. We are hopeful that exact, or nonadiabatic, calculations of the polarizabilities and hyperpolarizabilities of H_2^+ and D_2^+ will help reduce the remaining discrepancies and provide the first clear signature of retardation, or a "Casimir" force, within a Rydberg molecule.

While very little data exits with kHz resolution for H_2 Rydberg states attached to an anisotropic core ($N^+ > 0$) [12, 11], we can still provide some interesting studies of the vector interaction. One of the more revealing new predictions is that the vector hyperpolarizability β_v is inversely proportional to the rotational inertia of an anisotropic homonuclear diatomic ion. This fact confirms our physical picture of the vector interaction. As the distant Rydberg electron revolves around the core, it tries to "drag" the core polarization vector with it. Unfortunately, the relatively large moment of inertia of the molecule strongly resists this drag, thereby minimizing the "torque" effect. As a result, the dynamic vector interaction is tiny in molecular Rydberg states. An analysis of the energy shifts generated by the vector interaction and by relativistic retardation interactions in Rydberg states of ($\nu^+=0, N^+=1$) H_2 [19] demonstrates that they produce distinct patterns of energy level splittings that should be experimentally observable. This theoretical work, in combination with more sensitive experiments that are currently underway, should reveal the quantitative nature of both of these subtle interactions in Rydberg states of H_2.

CHAPTER 2

ADIABATIC DESCRIPTION OF RYDBERG SYSTEMS

In this chapter we present an adiabatic analysis of Rydberg systems that shows the origin of the vector interaction, and develop a multichannel formulation that generalizes perturbative polarization models. The goal here is to use techniques that are standard in atomic and molecular physics to derive a long range potential that describes how a slow, or adiabatic, Rydberg electron interacts with an arbitrary anisotropic ionic core.

2.1 Adiabatic Representation

The concept of adiabaticity has been used extensively in atomic and molecular physics. One of the earliest quantum mechanical application was the extremely useful Born-Oppenheimer approximation, which provides much of the framework used to understand the low vibrational states of diatomic molecules. The approximation essentially assumes that the relative motion of the nuclei is slow compared with the motion of the molecular electrons, and, as a consequence, the electronic properties of the system acquire a weak parametric dependence on the internuclear separation. In other words, the slow nuclear motion leads to an approximate separation of the nuclear and electronic degrees of freedom.

This approach can be extended to many problems in physics, provided there is at least one "slow" degree of freedom relative to all other degrees of freedom in the system. In Rydberg atoms and molecules it is the radial coordinate of the Rydberg electron that can serve as the adiabatic coordinate. A Rydberg atom or molecule is a simple quantum system involving an electron with high angular momentum ($\ell \geq 5$) and a relatively isolated ionic core. The slow moving Rydberg electron can be regarded as a probe of the electrostatic properties of the ionic core: specifically the permanent multipole moments and polarizabilities of the

relevant ionic states. The large spatial separation between high-ℓ Rydberg and core electronic states makes Rydberg electron spin effects (such as exchange) negligible. (These effects can be included perturbatively if they become appreciable.) The absence of exchange between the Rydberg and the core electrons leads to the $(J_c\ell)K$ coupling scheme [35, 36]. Here the total angular momentum J_c of the fast core electrons is coupled with the orbital momentum ℓ of the slow Rydberg electron to form a resultant K, which serves as the total angular momentum in most of the succeeding derivations.

The centrifugal repulsion associated with high-ℓ Rydberg states forces the Rydberg electron to roam beyond the confines of the core. In this distant region the radial coordinate of the Rydberg electron becomes distinct, both spatially and dynamically, from the coordinates of the ionic core. For these reasons, a useful expansion of the wave function for the entire system is

$$\Psi(r,\omega) = \sum_i \phi_i(\omega)\psi_i(r) \tag{2.1}$$

where the $\{\phi_i(\omega)\}$ representation is an r-independent basis set and ω represents all coordinates in the system except the Rydberg radial coordinate. This basis set is formed from a complete set of ionic energy eigenstates, whose angular momenta are coupled with the orbital functions Y_{lm} of the Rydberg electron. This *ansatz* for the wavefunction, combined with the Schrödinger equation, leads to a set of coupled radial equations that are usually called the "close-coupling equations" (without exchange) in atomic physics. The functions $\{\psi_i(r)\}$ in this primitive expansion can be viewed as the radial wavefunctions for motion within the ionic channels of the system. The potential curves that describe the radial motion of the Rydberg electron are formed from a combination of the diagonal matrix element $V_{ii}(r)$ of the Rydberg-core interaction potential, the repulsive centrifugal term, and the ionic threshold energy E_i in channel i. The off-diagonal matrix elements $V_{ji}(r)$ can then be viewed as causing Rydberg electron transitions from the $i-th$ to the $j-th$ Rydberg-core channel.

The full effective potential matrix that enters the conventional close-coupling (CC) equations without exchange is

$$V_{ij}^{CC}(r) = \left(\frac{\ell_i(\ell_i+1)}{2r^2} - \frac{1}{r} + E_i \right) \delta_{ij} + V_{ij}(r). \tag{2.2}$$

Here ℓ_i is the orbital momentum of the Rydberg electron in channel i. Unless stated otherwise, matrix elements involve integrals over all coordinates (ω) (and traces over all spins) in the problem, except for the radial coordinate of the Rydberg electron. Note that we have split off the dominant Coulomb interaction $-\frac{1}{r}$ so that the remaining potential matrix $V_{ij}(r)$ can be expanded as an asymptotic series in $\frac{1}{r}$.

2.2 Adiabatic Close-Coupling Equations

The primitive basis set $\{\phi_i(\omega)\}$ used in the conventional close-coupling equations is independent of the radial coordinate r of the Rydberg electron. As a consequence, this representation does not describe the radial variation in the polarization of the ionic core during close encounters with the Rydberg electron. That physics, however, is included in the standard close-coupling equations through off-diagonal channel coupling. Here we introduce a representation that builds the predominant dynamical effects of the Rydberg electron into a more useful effective potential. As in the molecular Born-Oppenheimer approach, these potential curves are obtained by diagonalizing an "adiabatic Hamiltonian" $\hat{H}_{r=const}$ in which derivative operators with respect to the adiabatic coordinate r are discarded. The resulting adiabatic eigenstates $\Phi_\mu(r;\omega)$ form (at every value of r) a complete orthonormal set in the coordinates ω. The adiabatic approximation is valid in our present context if the system remains confined within a single adiabatic channel as the Rydberg electron roams slowly beyond the confines of the core. Explicitly, the adiabatic potentials $U_\mu(r)$ and eigenstates $\Phi_\mu(r;\omega)$ are defined as r-dependent parametric solutions of the linear eigenvalue problem:

$$\hat{H}_{r=const}\Phi_\mu(r;\omega) = U_\mu(r)\Phi_\mu(r;\omega). \tag{2.3}$$

In the $\{\phi_i(\omega)\}$ representation, the matrix of the adiabatic Hamiltonian operator $\hat{H}_{r=const}$ reduces to $\underline{V}^{CC}(r)$. The adiabatic channel functions are r-dependent superpositions of the ionic core states and the orbital functions of the Rydberg electron. They contain information concerning the instantaneous interactions between the Rydberg electron and the core, and provided r varies slowly, concerning the approximately conserved properties of the electron-core system. An expansion of the total wavefunction for the system in terms of the adiabatic channel functions

$$\Psi(r,\omega) = \sum_\mu F_\mu(r)\Phi_\mu(r;\omega), \tag{2.4}$$

transforms the Schrödinger equation into a set of coupled radial equations. These can be written in matrix form as

$$\left[-\frac{1}{2}\left(\underline{I}\frac{d}{dr} + \underline{P}(r)\right)^2 - (E\underline{I} - \underline{U}(r))\right]\underline{F}(r) = 0, \tag{2.5}$$

where the derivative coupling matrix is defined by

$$P_{\mu\nu}(r) = \langle\Phi_\mu \mid \frac{\partial}{\partial r}\Phi_\nu\rangle. \tag{2.6}$$

Here we adopt Greek letters to label the adiabatic channels. At sufficiently large distances $r \to \infty$, each adiabatic channel converges to one of the ionic channels labeled by Roman letters in the primitive close-coupling representation.

In contrast to the primitive close-coupling representation, in which the coupling among the ionic channels enters through the potential matrix V^{CC}, the coupling among

our *adiabatic channels* derives from the derivative matrix $P_{\mu\nu}(r)$ that modifies the radial momentum operator. Thus the P-matrix accounts for all nonadiabatic or inelastic effects that arise because the slow Rydberg electron is not slow enough for the core electrons to instantaneously adjust within the time frame of Rydberg motion. For most of the regimes discussed here, the off-diagonal derivative couplings are small compared with the diagonal adiabatic potentials. Under these conditions, the motion of the system remains confined, to an excellent approximation, within a single adiabatic channel of the system. As we discuss later, the analogy between the modified radial momentum operator and a generalized momentum operator that involves both the usual mechanical momentum plus an additional term, such as an electromagnetic vector potential, [27, 37, 38] permits some of the long range interactions to be viewed as non-Abelian gauge fields.

2.3 Adiabatic Potentials and Post-Adiabatic Corrections

The Born-Oppenheimer approximation, which neglects the channel coupling matrix $\underline{P}(r)$ altogether, is adequate for many purposes. To achieve higher accuracy or to treat higher energy processes, however, it is important to include some effects of $\underline{P}(r)$. One natural approach to this problem attempts to write the radial Schrödinger equation in an approximate single channel form

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} - (E - u_\mu(r,E))\right] F_\mu(r) = 0, \tag{2.7}$$

in which a new effective potential $u_\mu(r,E)$ is introduced that depends on the adiabatic potentials $\{U_\nu(r)\}$, the derivative couplings, and the energy of the system E in some simple algebraic way.

The post-adiabatic theory of Klar and Fano[39, 40], later generalized by Aquilanti [41], provides a means of including the perturbative effects of derivative coupling through an iterative algebraic procedure. The main idea is to transform the original adiabatic equation

Eq.(2.5) into a set of coupled single-channel equations where the new coupling arises from terms proportional to $\underline{P}(r)$ or $\frac{d}{dr}\underline{P}(r)$. This "postadiabatic procedure" can be iterated, in principle, which will (hopefully) reduce the coupling strength in each successive iteration.

For our purposes we use only the first iteration of the procedure, for which the effective potential is given by

$$u_\mu(r,E) \simeq U_\mu(r) - \frac{1}{2}\left(P^2\right)_{\mu\mu} + 2\left(E - U_\mu(r)\right)\sum_\nu \frac{\mid P_{\mu\nu} \mid^2}{U_\mu(r) - U_\nu(r)}. \tag{2.8}$$

The perturbative condition $P^2_{\mu\nu}(r) \ll \mid U_\mu(r) - U_\nu(r) \mid$ should be satisfied at all r relevant in the problem, which should always be the case for sufficiently high ℓ states of Rydberg systems. An immediate implication of the nonadiabatic corrections is an increase in the potential energy due to the repulsive energy-independent diagonal term $-P^2_{\mu\mu}$. Inclusion of this term in molecular problems is often called the "adiabatic approximation" as opposed to the strict Born-Oppenheimer approximation which neglects $\underline{P}(r)$ completely. This term is usually written as a second derivative coupling matrix in that context, but it is equivalent to our form as the square of the first-derivative coupling matrix [39]. An additional implication is the presence of an energy-dependent contribution; its sign can vary from state to state, but for the channel of lowest energy, it generally makes the effective post-adiabatic potential $u_\mu(r,E)$ *increasingly attractive* as the energy *increases* above threshold.

The derivative couplings $P_{\mu\nu}(r)$ can be computed from the radial partial derivative of the adiabatic Hamiltonian. The diagonal derivative couplings vanish, that is $P_{\mu\mu}(r) = 0$, since the P matrix is skew symmetric, and the off diagonal terms are given by

$$P_{\mu\nu}(r) = \frac{\langle \Phi_\mu \mid \frac{\partial}{\partial r}\hat{H}_{r=const} \mid \Phi_\nu \rangle}{U_\nu(r) - U_\mu(r)}, \tag{2.9}$$

which can be derived from the defining equation, Eq.(2.3), for the adiabatic potentials and eigenfunctions. The partial derivative of the adiabatic Hamiltonian is readily evaluated

analytically. This form of the derivative coupling matrix clearly demonstrates the apparently singular behavior when two channels of common symmetry are nearly degenerate ($U_\mu(r) \approx U_\nu(r)$). The avoided crossing redistributes radial flux among the strongly coupled, nearly degenerate channels.

2.4 Perturbative Adiabatic Diagonalization for Nondegenerate Channels

In Rydberg atoms and molecules of high orbital momentum ($\ell \geq 5$) the dominant electrostatic interaction between the Rydberg electron and the ionic core is the dipole moment which goes as $\frac{1}{r^2}$. Since a typical inner turning radius is $r \approx \frac{\ell(\ell+1)}{2}$ for a Rydberg electron near zero energy, all electrostatic interactions are perturbative compared to the screened Coulomb attraction $-\frac{1}{r}$. The small values of the electrostatic matrix elements $V_{ij}(r)$ compared with the ionic threshold splittings validates a perturbative diagonalization of the $\underline{V}^{CC}(r)$ matrix.

An important step in a systematic perturbative diagonalization of this matrix is the inclusion of diagonal elements $V_{ii}(r)$ in the unperturbed Hamiltonian (see the *modified iteration-perturbation* formulas of Morse and Feshbach[42]). This is immediately apparent when the long range coupling matrix is written as

$$V_{ij}^{CC}(r) = \left(\frac{\ell_i(\ell_i+1)}{2r^2} - \frac{1}{r} + E_i + V_{ii}(r)\right)\delta_{ij} + V_{ij}(r), \tag{2.10}$$

where the $V_{ij}(r)$ are now purely off-diagonal contributions. A spherical multipole expansion of $V_{ij}(r)$ now separates the ionic core and Rydberg electron coordinates

$$\begin{aligned} V_{ij}(r) &= \sum_{k=0}^{\infty} \langle \phi_i \mid \frac{r_i^k}{r^{k+1}} P_k(cos\theta_{ir}) \mid \phi_j \rangle \\ &= \sum_{k=0}^{\infty} \langle \phi_i \mid \frac{r_i^k}{r^{k+1}} C^{(k)}(\hat{r}_i) \cdot C^{(k)}(\hat{r}) \mid \phi_j \rangle \end{aligned} \tag{2.11}$$

where the $C^{(k)}$ are renormalized spherical harmonics [43].

The perturbative diagonalization of the $\underline{V}^{CC}(r)$ matrix, through second order in $V_{ij}(r)$, produces terms involving summations over intermediate channels ν with potential energy denominators $U_\mu(r) - U_\nu(r)$. These second order contributions can be classified as either degenerate or nondegenerate depending on whether an intermediate channel ν is degenerate with the physically relevant channel μ at $r \to \infty$ (i.e., $U_\nu = U_\mu$). In the nondegenerate case (i.e., $E_\nu \neq E_\mu$), this approach gives a long range potential with the structure

$$u_\mu(r) = E_\mu - \frac{1}{r} + \frac{\ell_\mu(\ell_\mu+1)}{2r^2} + \frac{Q^{(2)}_{\mu\mu}}{r^3} - \frac{\alpha_\mu}{2r^4} + \frac{Q^{(4)}_{\mu\mu}}{r^5}$$

$$+ \frac{\beta^{ad}_\mu + \beta^{nad}_\mu - 2(E-E_\mu)\lambda_\mu - \delta_\mu - \eta_\mu}{2r^6} + O\left(\frac{1}{r^8}\right) \tag{2.12}$$

where *ad* (or *nad*) denotes adiabatic (or nonadiabatic). Every term in this potential can be written as a standard second-order perturbation sum, except for the diagonal quadrupole $Q^{(2)}_{\mu\mu}$ and hexadecapole $Q^{(4)}_{\mu\mu}$ terms which are diagonal (first-order) matrix elements of the ionic electric quadrupole and hexadecapole operators. Explicit expressions for these terms as infinite perturbation sums, over bound and continuum states of the core, can be obtained along the lines of the derivation given by Ref.[44], although there are differences in notation, in coupling scheme, and in the multipoles that were included. Each term $Q^{(2)}_{\mu\mu}$,$Q^{(4)}_{\mu\mu}$,α_μ, β^{ad}_μ, β^{nad}_μ, λ_μ, δ_μ, and η_μ depends on the various quantum numbers J_μ, ℓ_μ, and K in a relatively complicated fashion which is difficult to analyze:

$$Q^{(k)}_{\mu\nu} = \langle \mu \mid \sum_{i=1}^{N_e} r_i^k P_k(cos\theta_{ir}) \mid \nu \rangle, \tag{2.13}$$

$$\alpha_\mu = \sum_{\nu \neq \mu} \frac{2Q^{(1)}_{\mu\nu} Q^{(1)}_{\nu\mu}}{E_\nu - E_\mu}, \tag{2.14}$$

$$\beta^{ad}_\mu = \sum_{\nu \neq \mu} \frac{[\ell_\nu(\ell_\nu + 1) - \ell_\mu(\ell_\mu + 1)]}{(E_\nu - E_\mu)^2} Q^{(1)}_{\mu\nu} Q^{(1)}_{\nu\mu}, \tag{2.15}$$

$$\beta^{nad}_\mu = \sum_{\nu \neq \mu} \frac{4Q^{(1)}_{\mu\nu} Q^{(1)}_{\nu\mu}}{(E_\nu - E_\mu)^2}, \tag{2.16}$$

$$\lambda_\mu = \sum_{\nu \neq \mu} \frac{8Q^{(1)}_{\mu\nu} Q^{(1)}_{\nu\mu}}{(E_\nu - E_\mu)^3}, \tag{2.17}$$

$$\delta_\mu = \sum_{\nu \neq \mu} \frac{2Q^{(2)}_{\mu\nu} Q^{(2)}_{\nu\mu}}{E_\nu - E_\mu}, \tag{2.18}$$

and

$$\eta_\mu = \sum_{\nu \neq \mu} \frac{4Q^{(1)}_{\mu\nu} Q^{(3)}_{\nu\mu}}{E_\nu - E_\mu}. \tag{2.19}$$

In Eq. (2.13) N_c denotes the number of ionic core electrons, while the subscript r refers the Rydberg electron. One consequence of including nonadiabatic effects is the appearance of an energy dependent term in the long range potential. This energy dependence has generated a small controversy because different treatments disagree in multiplicative constants[21, 22]. Ref.[22] shows, however, that the energy dependent term of order $\frac{1}{r^6}$ can be written as an ℓ-dependent linear combination of $\frac{1}{r^7}$ and $\frac{1}{r^8}$, implying that the energy dependent term can be regarded as a contribution of higher order than $\frac{1}{r^6}$.

2.5 Degenerate Channel Contributions

The nature of the long range potential changes qualitatively when degenerate terms are present. For instance, in Mg the 3pnf and 3pnh channels are degenerate. Second order degenerate contributions appear when intermediate channels ν share the same threshold energies $E_\nu = E_\mu$, K-value and parity with the physically relevant channel μ at $r \to \infty$. However, for simplicity we assume that (as in the Mg case cited above) the intermediate Rydberg electron orbital momentum ℓ_ν differs from ℓ_μ, for all degenerate channels. A derivation similar to that given above produces two additional contributions to the potential $u_\mu(r)$ that originate in coupling to the degenerate channels:

$$u'_\mu(r) = u_\mu(r) + \frac{a_\mu}{r^4} + \frac{b_\mu}{r^6} + O\left(\frac{1}{r^8}\right) \tag{2.20}$$

where the a_μ and b_μ terms are explicitly

$$a_\mu = \sum_{\nu \neq \mu} \frac{2Q^{(2)}_{\mu\nu} Q^{(2)}_{\nu\mu}}{\ell_\mu(\ell_\mu + 1) - \ell_\nu(\ell_\nu + 1)} \tag{2.21}$$

and

$$b_\mu = \sum_{\nu \neq \mu} \frac{4Q^{(2)}_{\mu\nu} Q^{(4)}_{\nu\mu}}{\ell_\mu(\ell_\mu + 1) - \ell_\nu(\ell_\nu + 1)}. \tag{2.22}$$

Eq.(2.20) excludes one important type of degenerate channel coupling that arises when two or more ionic states of opposite parity are degenerate and can consequently support a permanent dipole moment. The resulting degenerate dipole-coupling of channels can be important, e.g. in the properties of doubly-excited states of the hydrogen negative ion. In this case a diagonalization within the degenerate channel space should be performed as in Refs.[45, 21]. After transformationing to a representation where the long-range dipole

coupling is diagonal, however, the effects of other (non-dipole) multipoles can be handled using the present techniques.

2.6 Advantages and Disadvantages

An adiabatic analysis is not quantitatively accurate in situations where curve crossings arise, because of the neglect of channel coupling. However, the qualitative insight provided by the single channel description can be a helpful tool when it comes to interpretation of the dynamics. For a Rydberg atom or molecule, the simple form of the long range adiabatic potential offers physical insight into the Rydberg-core interaction. Specifically, if the core of a Rydberg system is spherically symmetric (total angular momentum zero), then the interaction between the Rydberg electron and the many core electrons is purely scalar and the dominant interaction is induced dipole polarization. In anisotropic systems where the core is nonspherical (nonzero total angular momentum) the dominant interaction is the tensor quadrupole interaction followed by scalar and tensor dipole polarization. In chapter 4 we show that a new type of interaction arises when the core is anisotropic. The new vector interaction

$$V_{vector} = \beta_v \frac{\vec{L}_c \cdot \vec{\ell}}{r^6} \tag{2.23}$$

couples the angular momentum of the core with the orbital momentum of the Rydberg electron. The vector hyperpolarizability β_v depends on the dipole moments of the core and arises as a dynamic angular interaction due to the motion of the Rydberg electron.

CHAPTER 3

DIABATIC FORMULATION OF THE RYDBERG CHANNEL INTERACTIONS

In this chapter we present a multichannel diabatic formulation that provides an effective means of treating coupling among different Rydberg-core channels. By "eliminating" a class of channels, we use a Green's function expansion to derive an effective Hamiltonian with a potential matrix that can couple different μ and μ' channels. The resulting structure resembles that of the nondegenerate long range potential developed in the previous chapter.

3.1 Diabatic Representation

In the previous chapter the description of a Rydberg system began with the choice of a primitive representation $\{\phi_i(\omega)\}$ in the Hilbert space of the ion and of the angular degrees of freedom of the Rydberg electron. This basis set is independent of the radial coordinate r of the Rydberg electron. Through a diagonalization of the ionic Hamiltonian matrix (at every value of r) we form an adiabatic representation where each eigenstate is a superposition of the ionic core states and the orbital functions of the Rydberg electron. These adiabatic eigenstates contain information concerning the instantaneous Rydberg-core interaction and, provided channel coupling is weak, concerning the approximately conserved properties of the system. Unfortunately, an adiabatic analysis that involves a single adiabatic channel is usually only appropriate when channel coupling is weak, implying an approximate separability of the problem. In general, an accurate description of a Rydberg system requires the inclusion of channel coupling, either through the derivative coupling matrix elements $P_{\mu\nu}(r)$ of an adiabatic formulation or through the off-diagonal $V_{ij}(r)$ matrix elements in our primitive basis set expansion. Examples of this situation arise in our applications of these methods to Rydberg states of Ne[30, 31], to doubly excited autoionizing Rydberg states of

Mg[33], and finally to Rydberg states of H_2 and D_2 [34].

Our theoretical description of Rydberg systems is based on the primitive $\{\phi_i(\omega)\}$ basis set, which is often referred to as a "diabatic" representation. Adiabatic formulations and their corresponding potential curves are useful for developing a qualitative understanding of radial motion and for providing insight into perturbations and resonance structure. Unfortunately, the derivative coupling matrix elements of an adiabatic formulation can become nearly singular, when channels of common symmetry possess close avoided crossings. Such crossings occur ubiquitously in open shell atoms with fine structure and can lead to numerical difficulties; consequently, a diabatic formulation is often preferable for quantitative calculations.

The channel structure provided by the primitive close-coupling equations enables physically dominant channels to be separated from those that play only a perturbative role in the physics of a Rydberg system. Since most experiments devoted to the study of Rydberg energy levels focus on Rydberg series that converge to specific ionization thresholds of the core (typically split by fine structure or rotational interactions in the case of molecules), the total channel space can be partitioned into a P subspace of physically dominant channels and a complementary Q subspace (see Figure 3.1) that is needed to describe core multipole moments induced by the electric field of the Rydberg electron. In what follows, the interactions among channels belonging to the P subspace are treated "exactly", while couplings to the Q subspace are described perturbatively.

3.2 Diabatic Close-Coupling Equations

The diabatic close-coupling equations connecting all channels with specified total angular momentum K (excluding the Rydberg electron spin) and specified parity π can be written in the partitioned form

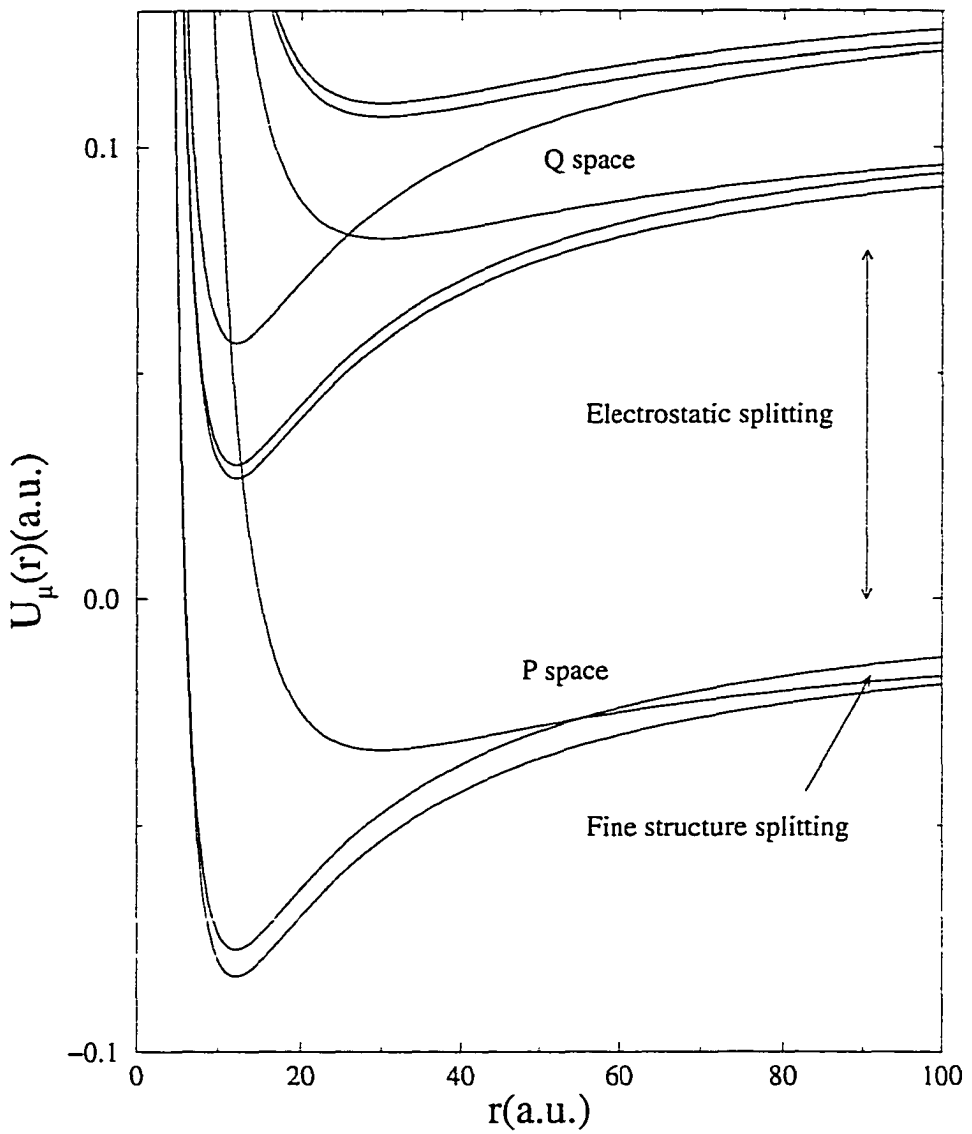


Figure 3.1. Generic radial potentials, with electrostatic (or vibrational) and fine-structure (or rotational) splittings, that demonstrate the natural separation of physically dominant channels from those that play only a perturbative role in the physics of a Rydberg system.

$$\left[-\frac{1}{2}\underline{I}\frac{d^2}{dr^2} - E\underline{I} + \begin{pmatrix} V^{QQ} & V^{QP} \\ V^{PQ} & V^{PP} \end{pmatrix}\right] \begin{pmatrix} \psi_Q \\ \psi_P \end{pmatrix} = 0, \tag{3.1}$$

where V is the coupling matrix V^{CC} in Eq.(2.2), and $\underline{I}$ is the identity matrix.

The number of these equations is generally infinite and for practical reasons must be truncated in some sensible way. The object is to derive a set of equations and an effective Hamiltonian that describe the motion of a Rydberg electron within the physical channel subspace. For this approach to be useful, the Hamiltonian must include coupling among the channels of the physical P subspace, and incorporate effects of the perturbative Q subspace channels, which manifest themselves through polarization and hyperpolarization interactions.

An equation describing the P subspace wavefunction and an effective Hamiltonian can be obtained through the standard technique of channel elimination [46, 47, 48]

$$H^{PP}\psi_P = E\psi_P, \tag{3.2}$$

using a Green's functions [49, 50]. The energy dependence of the effective Hamiltonian H^{PP}

$$H^{PP} = -\frac{1}{2}I^{PP}\frac{d^2}{dr^2} + V^{PP} + V^{PQ}G^{QQ}(E)V^{QP}, \tag{3.3}$$

is a consequence of the channel elimination, and enters through the radial Green's function matrix $G^{QQ}(E)$ that satisfies the equation

$$\left[-\frac{1}{2}I^{QQ}\frac{d^2}{dr^2} - EI^{QQ} + V^{QQ}\right] G^{QQ}(r, r', E) = -I^{QQ}\delta(r - r'), \tag{3.4}$$

subject to closed-channel boundary conditions. The dominant coupling among P subspace channels enters through the first order terms V^{PP}, while the perturbative effects of the Q subspace channels enters through the second-order terms involving the energy-dependent Green's function. The Green's function of equation (3.4) can be written in the operator form

$$G^{QQ}(r,r',E) = \left[E - H^{QQ}(r)\right]^{-1}\delta(r-r') \tag{3.5}$$

where the inverse operator $\left[E - H^{QQ}(r)\right]^{-1}$ is an integral operator with the boundary conditions built into it [49, 50]. In the next section we use a binomial expansion of this inverse operator to obtain an energy-independent effective Hamiltonian that can be used to accurately describe Rydberg motion within the physical channel subspace.

3.3 Generalized Eigenvalue Problem

The form of the effective Hamiltonian in equation (3.3) is neither very appealing nor useful, because of its nonlinear dependence on energy. Since we are interested in Rydberg states that converge to specific ionic thresholds of the core, this nonlinear energy dependence can be approximately removed through a binomial expansion of the Green's function and a simple energy transformation.

Channels in the P subspace are typically split by fine structure interactions (rotational interactions in the context of Rydberg molecules) while the energy difference between a channel in the P subspace and a channel in the Q subspace basically originates from electrostatic interactions (vibrational interactions). The difference in magnitude of fine structure and electrostatic splittings enables us to expand the Green's function in reciprocal powers of $E-E_Q$, where E_Q is an ionization threshold in the Q subspace. In addition to this expansion we also replace the energy of the system E with $E_{P_0}+\varepsilon$ where E_{P_0} is typically the ionic threshold within the P channel subspace under experimental study and ε is a small Rydberg energy shift. The resulting Green's function takes the simple, analytic form

$$G^{QQ}(r,r',E) = g^{QQ}(r,\varepsilon)\delta(r-r') \tag{3.6}$$

where

$$g^{QQ}(r,\varepsilon) \equiv \left[E - H^{QQ}(r)\right]^{-1} = \left[(E - E_Q) - (H^{QQ}(r) - E_Q)\right]^{-1}$$

$$= \frac{I^{QQ}}{E - E_Q} + \frac{I^{QQ}}{E - E_Q}(H^{QQ}(r) - E_Q)\frac{I^{QQ}}{E - E_Q} + \cdots$$

$$= \frac{I^{QQ}}{E_{P_0} - E_Q} + \frac{I^{QQ}}{E_{P_0} - E_Q}(H^{QQ}(r) - E_Q)\frac{I^{QQ}}{E_{P_0} - E_Q}$$

$$-\varepsilon\frac{I^{QQ}}{(E_{P_0} - E_Q)^2} + \cdots. \tag{3.7}$$

This formally exact analytic expression for the Green's function can be obtained, alternatively, through an eigenfunction expansion, for example, using the bound and continuum states of hydrogen [49, 50]. Since we are considering high-ℓ Rydberg states within the P subspace, the Green's function for the Rydberg electron can only couple to high-ℓ states of the Q subspace. The high-ℓ nature of this coupling suggests that only near threshold bound and continuum states within the Q subspace will contribute to the Green's function. In this sense, the $H^{QQ}(r) - E_Q$ operator represents a Rydberg energy that is small compared with the large electrostatic splittings. These arguments are supported by the fact that all dynamic or nonadiabatic effects arising from terms proportional to $\frac{1}{(E_{P_0}-E_Q)^2}$ are typically an order of magnitude smaller than the static or adiabatic terms proportional to $\frac{1}{E_{P_0}-E_Q}$. This will become clear in the later chapters where numerical calculations support these arguments.

The preceding analysis enables us to cast our original close-coupling equation, Eq.(3.2), involving an energy-dependent Hamiltonian, in the form of a generalized eigensystem (linear in ε) where the eigenvalue is the Rydberg energy ε

$$\bar{H}^{PP}\psi_P = \varepsilon\Lambda^{PP}\psi_P. \tag{3.8}$$

The energy-independent "effective Hamiltonian" $\bar{H}^{PP}$ is

$$\bar{H}^{PP} = -\frac{1}{2} I^{PP} \frac{d^2}{dr^2} - E_{P_0} I^{PP} + V^{PP} + V^{PQ} g^{QQ}(r,0) V^{QP}. \tag{3.9}$$

(Note: the $g^{QQ}(r,0)$ matrix in $\bar{H}^{PP}$ is evaluated at $\varepsilon = 0$, since we have collected all ε terms on the right-hand side of the generalized eigenvalue equation.) The Λ^{PP} matrix, whose presence in the eigenvalue equation is reminiscent of an overlap matrix in a nonorthogonal representation, is

$$\Lambda^{PP} = I^{PP} + V^{PQ} \frac{I^{QQ}}{(E_{P_0} - E_Q)^2} V^{QP}. \tag{3.10}$$

As in the adiabatic formulation, a spherical expansion of the electrostatic matrix V can be performed that separates all the core electronic and Rydberg coordinates. Such an expansion for the diabatic formulation is postponed until the next chapter where we present the recoupling theory and an effective parametrization of the long-range electron-core potential energy.

The interaction terms within the effective Hamiltonian and the Λ matrices are classified as permanent electric multipole , and induced "static" and "dynamic" polarization interactions (sometimes referred to as polarization and hyperpolarization). The static and dynamic polarization interactions comprise all second-order contributions, and are distinguished from one another by their dependence on information regarding the motion of the Rydberg electron. Specifically, *dynamic* terms involve either radial or angular kinetic energy operators of the Rydberg electron, while *static* terms depend solely upon the electrostatic interaction between the Rydberg electron and the ionic core. The dynamic terms can be viewed as representing coupling due to the *motion* of the Rydberg electron.

In principle, the expansion of the Green's function can be carried out to higher-order in ε. Unfortunately, such an expansion only complicates the description of Rydberg electron motion by introducing higher-order nonlinear energy dependence with no significant improvement in energy level nor wavefunction information. Another point deserving attention is the connection between our approach and that of others like W.G. Sturrus [12, 11]

and W.G. Schoenfeld [23]. While our derivation differs from their theoretical analysis, the final effective Hamiltonians exhibit similarities. The notable difference between our approach and that of Sturrus is our nonperturbative inclusion of coupling among channels, and consequently among different Rydberg series. In the approach of Sturrus, the effect of the core on the positions of Rydberg levels is incorporated through diagonal matrix elements of the effective Hamiltonian within a hydrogenic basis; this is tantamount to traditional second-order Rayleigh-Schrödinger perturbation theory. The coupling among different $n\ell$ states is included perturbatively, which breaks down when there are near-degeneracies among Rydberg states. In our formulation the coupling among Rydberg levels converging to thresholds within our physical channel subspace is treated to all orders by diagonalizing the generalized eigensystem. Sturrus' evaluation of dynamic corrections relies upon the approximation that ε can be replaced with the effective radial hydrogenic Hamiltonian $\varepsilon \rightarrow -\frac{1}{2}\frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{2r^2} - \frac{1}{r}$. By collecting all terms involving ε on the right hand side of our coupled equations, no such approximation is needed and all coupling among channels can be effectively included. This permits a more realistic description of Rydberg electron motion in the presence of an arbitrary ionic core. It might be noted that this type of rearrangement and energy linearization can also be applied to the post-adiabatic formulation of Klar and Fano [39, 40].

3.4 Autoionization Rates

In anticipation of Chapter 6, where we discuss autoionization in doubly excited Rydberg systems, we present a final section on the calculation of autoionization rates. The excited electrons in a doubly excited Rydberg system may exchange energy through their electron-electron interaction, forcing one or more core electrons to drop into a lower state, while ejecting the remaining electron into the continuum. (This does not involve the emission of electromagnetic radiation.) The total initial state wavefunction for the Rydberg system involves both $|\psi_P\rangle$ and $|\psi_Q\rangle$, components of the wavefunction distributed among all channels. Although $|\psi_P\rangle$ represents the dominant portion of the total wavefunction, the perturbative

contribution $|\psi_Q\rangle$ should be included for consistency and to ensure complete convergence in the calculation of decay rates.

The rate for a perturbative transition from the initial state $|\psi\rangle$ to a final energy normalized state $|\phi\rangle$, associated with a channel energetically below the P subspace, is given by the "Fermi Golden Rule"

$$\Gamma = 2\pi \mid \langle \phi \mid V \mid \psi \rangle \mid^2, \tag{3.11}$$

where V is the electrostatic interaction minus the screening $\frac{1}{r_>}$ potential. Expressing $|\psi\rangle$ in terms of P and Q subspace components this rate can be written as

$$\Gamma = 2\pi \mid \langle \phi \mid T^{QP} \mid \psi_P \rangle \mid^2, \tag{3.12}$$

where the transition matrix T^{QP} is

$$T^{QP} = V^{QP} + V^{QQ} g^{QQ} V^{QP}. \tag{3.13}$$

This final expression for the autoionization rate Γ shows that once $|\psi_P\rangle$ is known, the perturbative effects of $|\psi_Q\rangle$ can be included through use of the Green's function developed in the previous section.

3.5 Advantages

The primary advantage of a diabatic formulation over an adiabatic formulation is the ease with which channel coupling can be included. For many systems, channel coupling is essential for accuracy. The polarization models introduced by Sturrus [11, 12] and Schoenfeld [23] are attractive from a physical standpoint, but they have difficulty achieving the spectroscopic accuracy needed to compare theory with experiment. On the other hand,

the fact that electric multipole and induced polarization interactions are involved in these theories offers a simple physical picture of Rydberg and core electron interaction. The developments presented in the next chapter enable us to recast our multichannel theory in terms of multipole and polarization properties of the ionic core. However, our retention of channel coupling allows us to utilize both the physical insight of polarization interactions and the spectroscopic accuracy of a multichannel approach.

CHAPTER 4

RECOUPLING AND PARAMETRIZATION

In this chapter we disentangle the operator structure and the anisotropic nature of the long range potentials developed in the two previous chapters. The analysis presented here enables us to systematically derive, for the first time, the complete vector interaction, and to explain its physical origin and implications.

4.1 Motivation

The forms of the effective adiabatic potential developed in Chapter 2 and the effective diabatic Hamiltonian developed in Chapter 3 are simple from an algebraic standpoint, but each term in them depends on the various quantum numbers J_c, ℓ, and K in a relatively complicated fashion that is difficult to analyze. The objective in this chapter is to reveal the operator structure and to clarify the dependence of each interaction term on the various core and Rydberg quantum numbers. Once this recoupling is accomplished the concept of parametrization, which is the identification of channel dependent core properties, can be developed.

Each term within the adiabatic potential and diabatic Hamiltonian can be written in terms of spherical tensor operators that act on states of the core and the Rydberg electron. Using standard Wigner-Racah recoupling algebra these terms can be recoupled so that information pertaining to the core is separated from information pertaining to the Rydberg electron. As an example, the adiabatic dipole polarizability α_μ in Eq.(2.14) is proportional to the expectation value, in $|\mu\rangle$, of

$$(r_c^{(1)} \cdot r_r^{(1)}) P (r'^{(1)}_c \cdot r'^{(1)}_r) \tag{4.1}$$

where P is a weighted (scalar) projection operator. This expression can be recoupled into

$$\sum_k (-1)^k \left[r_c^{(1)} \otimes P_c r'^{(1)}_c \right]^{(k)} \cdot \left[r_r^{(1)} \otimes P_r r'^{(1)}_r \right]^{(k)}, \tag{4.2}$$

where factors appear with net multipole moment k acting on the core and Rydberg electron, respectively. The validity of this derivation relies on the fact that the infinite summation over intermediate states ν is itself a "scalar" object that contributes no multipolarity to any term. The expectation value in $|\mu\rangle$ of this expression can be written as

$$\alpha_\mu = \alpha_s^\mu + \alpha_t^\mu \mathcal{A}_{\mu\mu}^{(2)} \tag{4.3}$$

where α_s^μ and α_t^μ are the scalar and tensor dipole polarizabilities of the core in the adiabatic channel μ, and $\mathcal{A}_{\mu\mu}^{(2)}$ is an angular coupling factor, which is the mean value of a second-rank tensorial operator.

In the next few sections we demonstrate how a similar analysis can be extended to all interaction terms in the adiabatic potential and the diabatic Hamiltonian. Just as the Wigner-Eckart theorem enables the factorization of angular matrix elements into purely geometric and dynamic contributions, the tensorial analysis presented here enables a factorization of the bulk ionic properties of the core from the dynamic coupling of the Rydberg and core electronic interaction. Moreover, the scalar and tensor parameters, which represent bulk properties of the ionic core, can be computed from first principles, and indirectly extracted from experimental Rydberg spectra, thereby providing an important test of theoretical core wavefunction information.

4.2 Recoupling of Spherical Tensor Operators

Spherical tensor operators and the Wigner-Racah algebra used to manipulate them play a central role in atomic and molecular physics. A spherical tensor operator of rank k is defined to be a set of $2k+1$ functions $T_q^{(k)}$ with components $q = -k, -k+1, \cdots, k-1, k$ that transform under a coordinate frame rotation as

$$\underline{R} T_q^{(k)} \underline{R}^{-1} = \sum_p T_p^{(k)} D_{pq}^k(R) \tag{4.4}$$

where the expansion coefficients are the Wigner rotation matrix elements $D_{pq}^k(R)$ and R is a set of Euler angles.

The primary reason for introducing spherical tensor operators is that they greatly simplify the evaluation of angular matrix elements through the use of the Wigner-Eckart theorem. The Wigner-Eckart theorem disentangles matrix elements into a product of two factors; one that is purely geometric, expressing the symmetry and selection rules of the system, and another that contains the dynamics. The Wigner-Eckart theorem reads:

$$\langle \gamma j m \mid T_q^{(k)} \mid \gamma' j' m' \rangle = (-1)^{j-m} \begin{pmatrix} j & k & j' \\ -m & q & m' \end{pmatrix} \langle \gamma j \parallel T^{(k)} \parallel \gamma' j' \rangle, \tag{4.5}$$

where the phase and normalization conventions are those of Refs.[43, 51, 52]. Since the reduced matrix element does not depend of the magnetic quantum numbers m, m', and q it can be found once and for all by evaluating the full matrix element for a specific set of $\{m, m', q\}$, provided the matrix element is nonzero.

A key spherical tensor operator for our analysis is the renormalized spherical harmonic $C_q^{(k)}$

$$C_q^{(k)} = \left(\frac{4\pi}{2k+1}\right)^{\frac{1}{2}} Y_{k,q}(\theta, \phi) \tag{4.6}$$

where $Y_{k,q}(\theta,\phi)$ is the standard spherical harmonic function. The reduced matrix element of $C_q^{(k)}$ is [43]

$$\langle \ell \parallel C^{(k)} \parallel \ell' \rangle = (-1)^{\ell}[(2l+1)(2\ell'+1)]^{\frac{1}{2}} \begin{pmatrix} \ell & k & \ell' \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.7}$$

Moreover, the spherical harmonic addition theorem, which is used in the spherical expansion of $\frac{1}{r_{ij}}$, is written in this notation as

$$\begin{aligned} P_k(cos\theta_{ij}) \quad &= C^{(k)}(\hat{r}_i) \cdot C^{(k)}(\hat{r}_j) \\ &\equiv \sum_q (-1)^q C_q^{(k)}(\hat{r}_i) C_{-q}^{(k)}(\hat{r}_j), \end{aligned} \tag{4.8}$$

as defined in Ref.[43].

4.2.1 Recoupling of First-Order Terms The general form of all first-order terms in our perturbative expansion of the Rydberg-core interaction potential is

$$W^1(k) = A_c^{(k)} \cdot B_r^{(k)} \tag{4.9}$$

where $A_c^{(k)}$ and $B_r^{(k)}$ are spherical operators of rank k that act on the core and the Rydberg electron respectively. As is clearly seen these first-order operators are already written in a factorized form, making the evaluation of matrix elements rather trivial. The angular matrix element of this operator expression is [43]

$$\begin{aligned} W^1(k)_{\mu\mu'} \quad &= \langle \mu \mid A_c^{(k)} \cdot B_r^{(k)} \mid \mu' \rangle = \langle \gamma(J_c \ell) K M \mid A_c^{(k)} \cdot B_r^{(k)} \mid \gamma'(J_c' \ell') K' M' \rangle \\ &= \delta_{K,K'} \delta_{M,M'} (-1)^{J_c' + \ell + K} \begin{Bmatrix} J_c & \ell & K \\ \ell' & J_c' & k \end{Bmatrix} \langle \gamma J_c \parallel A_c^{(k)} \parallel \gamma' J_c' \rangle \langle \ell \parallel B_r^{(k)} \parallel \ell' \rangle. \end{aligned}$$

(4.10)

4.2.2 Recoupling of Second-Order Terms The general form of all second-order terms in our perturbative expansion of the Rydberg-core interaction is

$$W^2(k,k') = (T_c^{(k)} \cdot T_r^{(k)}) S^{(0)} (T_c^{(k')} \cdot T_r^{(k')}) \tag{4.11}$$

where $S^{(0)}$ is a weighted scalar projection operator that includes the energy denominators, *e.g.* of Eqs.(2.14) through (2.19). For a Rydberg system, this scalar operator can be written as a summation of projection operators, where each projection operator is product of two operators: one that projects onto the angular states of the Rydberg electron, and another that involves the energy denominators and projections onto the states of the core,

$$S^{(0)} = P_r P_c. \tag{4.12}$$

Then $W^2(k,k')$ can be rewritten as

$$W^2(k,k') = \left[T_c^{(k)} \cdot (T_r^{(k)} P_r)\right] \left[(P_c T_c^{(k')}) \cdot T_r^{(k')}\right] \equiv (A_c^{(k)} \cdot B_r^{(k)})(C_c^{(k')} \cdot D_r^{(k')}). \tag{4.13}$$

or in tensorial coupling notation as [53, 43, 51]

$$W^2(k,k') = (-1)^{k+k'} \left[(2k+1)(2k'+1)\right]^{\frac{1}{2}} \left[\left[A_c^{(k)} \otimes B_r^{(k)}\right]^{(0)} \otimes \left[C_c^{(k')} \otimes D_r^{(k')}\right]^{(0)}\right]_0^{(0)}. \tag{4.14}$$

Finally, a recoupling of the tensor products leads to the following form for the matrix element

$$W^2(k,k')_{\mu\mu'} = \langle \gamma(J_c \ell) K M \mid W^2(k,k') \mid \gamma'(J_c' \ell') K' M' \rangle$$

$$= (-1)^{k+k'} \sum_q (-1)^q \langle \mu \mid \left[A_c^{(k)} \otimes C_c^{(k')} \right]^{(q)} \cdot \left[B_r^{(k)} \otimes D_r^{(k')} \right]^{(q)} \mid \mu' \rangle$$

$$= (-1)^{k+k'+J_c'+\ell+K} \delta_{K,K'} \delta_{M,M'} \sum_q (-1)^q \begin{Bmatrix} J_c & \ell & K \\ \ell' & J_c' & q \end{Bmatrix}$$

$$\langle \gamma J_c \parallel \left[A_c^{(k)} \otimes C_c^{(k')} \right]^{(q)} \parallel \gamma' J_c' \rangle \langle \ell \parallel \left[B_r^{(k)} \otimes D_r^{(k')} \right]^{(q)} \parallel \ell' \rangle. \qquad (4.15)$$

4.2.3 Unit Tensor Notation In this subsection we introduce a unit tensor notation that allows us to write both $W^1(k)_{\mu\mu'}$ and $W^2(k,k')_{\mu\mu'}$ in terms of a common set of unit tensor operators. Let us define $X^{(k)}$ and $Y^{(k)}$ as "unit tensor operators" (spatial) of rank k that act on the ionic core and the Rydberg electron, respectively, such that

$$\langle X^{(k)} \cdot Y^{(k)} \rangle_{\mu\mu'} = \langle (J_c \ell) K \mid X^{(k)} \cdot Y^{(k)} \mid (J_c' \ell') K \rangle$$

$$\equiv \delta_{K,K'} \delta_{M,M'} (-1)^{J_c'+\ell+K} \langle J_c \parallel X^{(k)} \parallel J' \rangle \langle \ell \parallel Y^{(k)} \parallel \ell' \rangle \begin{Bmatrix} J_c & \ell & K \\ \ell' & J_c' & k \end{Bmatrix}$$

(4.16)

with the reduced matrix elements [54]

$$\langle J_c \parallel X^{(k)} \parallel J_c' \rangle = \begin{Bmatrix} 1, & if\ \Delta(J_c, k, J_c') \\ 0, & otherwise \end{Bmatrix} \qquad (4.17)$$

and

$$\langle \ell \parallel Y^{(k)} \parallel \ell' \rangle = \begin{Bmatrix} 1, & if\ \Delta(\ell, k, \ell') \\ 0, & otherwise \end{Bmatrix}. \qquad (4.18)$$

Here $\Delta(x,y,z)$ denotes the condition of triangularity among the quantum numbers $\{x,y,z\}$, *i.e.* Δ ="true" if $\mid x-y \mid \leq z \leq x+y$, and if $x+y+z$ =integer.

In this notation the first-order terms $W^1(k)_{\mu\mu'}$ in our perturbative expansion of the Rydberg-core interaction become

$$W^1(k)_{\mu\mu'} = \langle \gamma J_c \parallel A_c^{(k)} \parallel \gamma' J_c' \rangle \langle \ell \parallel B_r^{(k)} \parallel \ell' \rangle \langle X^{(k)} \cdot Y^{(k)} \rangle_{\mu\mu'}. \tag{4.19}$$

Likewise, the second-order terms $W^2(k,k')_{\mu\mu'}$ are

$$W^2(k,k')_{\mu\mu'} = \sum_q (-1)^{k+k'+q} \langle \gamma J_c \parallel \left[A_c^{(k)} \otimes C_c^{(k')} \right]^{(q)} \parallel \gamma' J_c' \rangle \langle \ell \parallel \left[B_r^{(k)} \otimes D_r^{(k')} \right]^{(q)} \parallel \ell' \rangle$$

$$\times \langle X^{(q)} \cdot Y^{(q)} \rangle_{\mu\mu'}. \tag{4.20}$$

The unit tensor notation helps elucidate the separation of each interaction into a product of three factors: one factor that depends on properties of the ionic core, another term that relates only to the Rydberg electron, and finally a simple angular factor that couples the angular momentum properties of the core with the orbital angular momentum of the Rydberg electron. It is important to note that no approximation has been introduced as a result of the tensorial analysis presented here. In the following sections we show that in special circumstances these unit tensor operators can be replaced with angular momentum operators that generate the same coupling.

4.3 Parametrization of Interactions

At this stage all first and second-order interaction terms in the effective adiabatic potential and diabatic Hamiltonian have been recoupled which factors out information pertaining to the ionic core from information relating to the Rydberg electron. However, the matrix elements that depend on core wavefunction information are still channel dependent. The channel dependence of these interactions is not strictly a problem, since in more traditional approaches like multichannel quantum defect theory and R-matrix methods [55, 56],

a multichannel reactance-matrix or scattering-matrix is formed that provides essentially the same information. Unfortunately, it is difficult to extract physical insight from large basis set and R-matrix calculations.

The tensorial analysis of the preceding section facilitates a parametrization of each interaction. Specifically, the ionic core properties of the first-order terms in our perturbative expansion of the Rydberg-core interaction are simply the permanent multipole moments of the ionic core. Similarly, the ionic properties of the second-order interaction terms can be identified as the channel dependent polarizabilities and hyperpolarizabilities that characterize the multipole moments of the core induced by the Rydberg electron.

For some systems, further approximations make it possible to approximately remove the channel dependence of each parameter. In atoms with low-Z atomic nuclear charge, the spin-orbit interaction is small compared with electrostatic interactions. Therefore the total orbital angular momentum L_c and the total spin S_c of the core are approximately good quantum numbers. For such light systems each parameter can be replaced with an $LS-$coupled version so that channel dependence arises only from the simple angular coupling factor.

4.4 Tensor Analysis of Adiabatic Potential

In the adiabatic theory of chapter 2 we developed both nondegenerate and degenerate contributions to the long range adiabatic potential that describes the interaction of a slow charge with an aribitrary anisotropic core. Each term in the adiabatic potential depends on quantum numbers of both the ionic core and the Rydberg electron in a rather complicated way. To clarify the dependence of each term on J_c, ℓ, and K, and to reveal the operator structure of the full potential, we apply the tensor recoupling theory of the previous sections so that each term takes the form

$$\sum_k (-1)^k c_k \langle X^{(k)} \cdot Y^{(k)} \rangle_\mu, \tag{4.21}$$

where $X^{(k)}$ and $Y^{(k)}$ are the unit tensor operators defined in subsection 4.2.3 that operate on the core and the Rydberg electron, respectively. Next, because of the single channel nature of the adiabatic potential we follow the spirit of the Fano-Macek [53] treatment of alignment and orientation and replace the tensorial structure by coupled angular-momentum operators of the same rank. Each such replacement introduces a compensating ratio of reduced matrix elements

$$\sum_k (-1)^k \frac{\langle \mu \parallel X^{(k)} \cdot Y^{(k)} \parallel \mu \rangle}{\langle \mu \parallel J_c^{(k)} \cdot \ell^{(k)} \parallel \mu \rangle} \langle J_c^{(k)} \cdot \ell^{(k)} \rangle_\mu = \sum_k C_k \langle J_c^{(k)} \cdot \ell^{(k)} \rangle_\mu, \tag{4.22}$$

where the choice of the $J_c^{(k)} \cdot \ell^{(k)}$ operator representation is motivated by the fact that adiabatic potentials involve specific values of J_c and ℓ_μ. In general, the choice of a particular operator representation depends on whether the formulation is adiabatic or diabatic. In our diabatic formulation the "unit tensor operator" notation is appropriate, and only under special circumstances, as in the case of low-Z atomic systems, can the unit operators be replaced with simpler angular momentum operators of the same rank that generate the same coupling among the Rydberg channels and reflect the basic symmetry properties of the system.

4.4.1 Nondegenerate Channel Contributions Keeping powers of $\frac{1}{r}$ up to radial order $\frac{1}{r^6}$ and grouping terms of the same tensorial structure allows us to present the nondegenerate long-range adiabatic potential in a form that better displays its anisotropic nature

$$u_\mu(r) = E_\mu - \frac{1}{r} + \frac{\ell_\mu(\ell_\mu+1)}{2r^2} + \frac{C^{4\mu}_{0(1,1)}}{r^4} + \frac{C^{6\mu}_{0[(1,1),(2,2)]}}{r^6} + \frac{C^{6\mu}_{1(1,1)}}{r^6} \langle J_c^{(1)} \cdot \ell^{(1)} \rangle_\mu$$

$$+\left(\frac{C^{3\mu}_{2(2,0)}}{r^3}+\frac{C^{4\mu}_{2(1,1)}}{r^4}+\frac{C^{6\mu}_{2[(1,1),(2,2),(1,3)]}}{r^6}\right)\langle J_c^{(2)}\cdot\ell^{(2)}\rangle_\mu$$

$$+\left(\frac{C^{5\mu}_{4(4,0)}}{r^5}+\frac{C^{6\mu}_{4[(2,2),(1,3)]}}{r^6}\right)\langle J_c^{(4)}\cdot\ell^{(4)}\rangle_\mu. \tag{4.23}$$

Here the terms $C^{n\mu}_{k(a,b)}$, corresponding to *even* tensorial rank k, of radial order $\left(\frac{1}{r}\right)^n$, and formed from the electric a-multipole and b-multipole moment contributions from $\frac{1}{r_{ij}}$, are given by

$$C^{n\mu}_{k(a,b)}=\frac{(-1)^{2J_c+\ell_\mu}(2\ell_\mu+1)(2k+1)}{\langle\ell_\mu\parallel\ell^{(k)}\parallel\ell_\mu\rangle\langle J_c\parallel J_c^{(k)}\parallel J_c\rangle}\begin{pmatrix}\ell_\mu & \ell_\mu & k\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}a & b & k\\ 0 & 0 & 0\end{pmatrix}$$

$$\sum_{\gamma_\nu J_\nu}\Gamma^n_k(a,b)\begin{Bmatrix}a & b & k\\ J_c & J_c & J_\nu\end{Bmatrix}\langle\gamma_cJ_c\parallel\sum_{i=1}^{N_c}r_i^aC^{(a)}(\hat{r}_i)\parallel\gamma_\nu J_\nu\rangle$$

$$\langle\gamma_\nu J_\nu\parallel\sum_{j=1}^{N_c}r_j^bC^{(b)}(\hat{r}_j)\parallel\gamma_cJ_c\rangle \tag{4.24}$$

where $C^{(a)}(\hat{r}_i)$ are renormalized spherical harmonics. Terms that share the same tensorial rank and power of $\frac{1}{r}$ but differ in multipole dependence are combined additively into a single net coefficient:

$$C^{n\mu}_{k[(a_1,b_1),(a_2,b_2),\cdots]}=C^{n\mu}_{k(a_1,b_1)}+C^{n\mu}_{k(a_2,b_2)}+\cdots. \tag{4.25}$$

The matrix elements $\langle J_c^{(k)}\cdot\ell^{(k)}\rangle_\mu$ are

$$\langle J_c^{(k)}\cdot\ell^{(k)}\rangle_\mu = \langle(J_c\ell_\mu)K\mid J_c^{(k)}\cdot\ell^{(k)}\mid(J_c\ell_\mu)K\rangle$$

$$=(-1)^{J_c+\ell_\mu+K}\langle J_c\parallel J_c^{(k)}\parallel J_c\rangle\langle\ell_\mu\parallel\ell^{(k)}\parallel\ell_\mu\rangle\begin{Bmatrix}J_c & \ell_\mu & K\\ \ell_\mu & J_c & k\end{Bmatrix}, \tag{4.26}$$

Ref.[43, 51], and explicit expressions for particular $\Gamma^n_k(a,b)$'s with even k in Eq. (4.24) are given by

$$\Gamma^3_k(2,0) = \Gamma^5_k(4,0) = 1, \tag{4.27}$$

$$\Gamma^4_k(1,1) = \Gamma^6_k(2,2) = \frac{1}{E_\mu - E_\nu}, \tag{4.28}$$

$$\Gamma^6_k(1,3) = 2\Gamma^4_k(1,1), \tag{4.29}$$

$$\Gamma^6_0(1,1) = \frac{6}{2(E_\nu - E_\mu)^2} - \frac{8(E - E_\mu)}{(E_\nu - E_\mu)^3}, \tag{4.30}$$

and

$$\Gamma^6_2(1,1) = \frac{3}{2(E_\nu - E_\mu)^2} - \frac{8(E - E_\mu)}{(E_\nu - E_\mu)^3}. \tag{4.31}$$

In contrast to the long-range potential presented in Eq.(2.12), the operator form of the anisotropic potential in Eq.(4.23) possesses a very simple dependence on the various quantum numbers J_c, ℓ, and K. Along with the factorization of the orbital angular momentum ℓ of the Rydberg electron from information pertaining to the ionic core, all terms of the same *even* tensorial rank k share the same ℓ-dependence. In addition, all dependence on K appears in a single $6-j$ symbol originating from the matrix element $\langle J_c^{(k)} \cdot \ell^{(k)} \rangle_\mu$, and accounts for the splitting of the $\mid K + J_c \mid - \mid K - J_c \mid +1$ number of ℓ levels of common K. Unfortunately, the terms $C^{n\mu}_{k(a_1,b_1)}, C^{n\mu}_{k(a_2,b_2)}, \cdots$ that make up $C^{n\mu}_{k[(a_1,b_1),(a_2,b_2),\cdots]}$ are not distinguishable from one another since they share the same tensorial rank, power of $\frac{1}{r}$, and ℓ-dependence.

The angular momentum representation of the nondegenerate long-range potential in Eq.(4.23) immediately shows the appearance of a vector contribution whose structure is

similar to a term predicted by Zygelman[27]. And, as noted in the Introduction, a term of this tensorial structure was introduced into atomic spectroscopy by Trees [28] and Racah [29], but on purely semiempirical grounds, with no explicit derivation or formal justification. As a complement and extension of the work of Zygelman, we provide an explicit expression for the coefficient of this unusual interaction and interpret its physical origin. The term $C^{6\mu}_{1(1,1)}$ is the *only* one, out to radial order $\frac{1}{r^6}$, with *odd* tensorial rank k ($k = 1$)

$$C^{6\mu}_{1(1,1)} = (-1)^{2J_c}\frac{\sqrt{6}}{\langle J_c \parallel J_c^{(1)} \parallel J_c\rangle}\sum_{\gamma_\nu J_\nu}\frac{1}{2(E_\nu - E_\mu)^2}\begin{Bmatrix} 1 & 1 & 1 \\ J_c & J_c & J_\nu \end{Bmatrix}$$

$$\langle \gamma_c J_c \parallel \sum_{i=1}^{N_e} r_i C^{(1)}(\hat{r}_i) \parallel \gamma_\nu J_\nu\rangle\langle \gamma_\nu J_\nu \parallel \sum_{j=1}^{N_e} r_j C^{(1)}(\hat{r}_j) \parallel \gamma_c J_c\rangle. \tag{4.32}$$

The existence of this term has been controversial [10, 11] since a simple parity argument might seem to negate the existence of any odd-rank tensor interaction in the long-range potential. The parity argument, however, fails since the "vector" interaction is an *even parity pseudovector* in the ionic core degrees of freedom. Surprisingly, this interaction would vanish if it were not for the $\ell_\nu(\ell_\nu + 1)$ factor in the summation for β^{ad}_μ, which allows the vector coefficient $C^{6\mu}_{1(1,1)}$ to survive.

Next we demonstrate that the untransformed potential, in Eq.(2.12), cannot support odd tensorial contributions, such as $J_c^{(1)} \cdot \ell^{(1)}$, without such an additional ℓ dependence. The ℓ_ν summation can be evaluated explicitly

$$\sum_{\ell_\nu}(-1)^{\ell_\nu}(2\ell_\nu + 1)\begin{pmatrix} \ell_\mu & a & \ell_\nu \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \ell_\nu & b & \ell_\mu \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} a & b & k \\ \ell_\mu & \ell_\mu & \ell_\nu \end{Bmatrix}$$

$$= (-1)^{a+b-k}\begin{pmatrix} a & b & k \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \ell_\mu & \ell_\mu & k \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.33}$$

Since each contribution to our long-range potentials involves a and b multipole moments that add up to an even number the tensorial rank k must be even in this case.

However, if there is an addition ℓ_ν dependence, such as $\ell_\nu(\ell_\nu+1)$ in β_μ^{ad}, the summation becomes

$$\sum_{\ell_\nu}(-1)^{\ell_\nu}\ell_\nu(\ell_\nu+1)(2\ell_\nu+1)\begin{pmatrix}\ell_\mu & 1 & \ell_\nu\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell_\nu & 1 & \ell_\mu\\ 0 & 0 & 0\end{pmatrix}\begin{Bmatrix}1 & 1 & 1\\ \ell_\mu & \ell_\mu & \ell_\nu\end{Bmatrix}$$

$$=(-1)^{\ell_\mu+1}\left[\frac{2}{3}\frac{\ell_\mu(\ell_\mu+1)}{(2\ell_\mu+1)}\right]^{\frac{1}{2}} \tag{4.34}$$

which is nonzero for Rydberg states with $\ell_\mu \neq 0$. Thus it is the centrifugal energy of the Rydberg electron, or angular motion, that generates the vector interaction.

4.4.2 Degenerate Channel Contributions The nature of the long range adiabatic potential changes qualitatively when second order degenerate contributions appear from intermediate channels ν that share the same ionization thresholds, K-value and parity with the physically relevant channel μ at $r \to \infty$. For example, the $3pnf$ and $3pnh$ channels in Mg are degenerate. However, as detailed in section 2.5, we assume that the intermediate Rydberg electron orbital momentum ℓ_ν differs from ℓ_μ, for all degenerate channels. The additional degenerate contributions to the long range adiabatic potential can be analyzed, as in the preceding subsection, to reveal their tensorial structure, which can then be replaced with coupled angular momentum operators of the same rank. Keeping powers of $\frac{1}{r}$ up to radial order $\frac{1}{r^6}$ and grouping terms of the same tensorial rank allows us to present these terms in the long range adiabatic potential that derive from the degenerate channels. The resulting degenerate terms, which are understood to be added to the nondegenerate potential of Eq.(2.12), are:

$$U_\mu^{degen}(r) = \frac{D^{4\mu}_{0(2,2)}}{r^4} + \frac{D^{4\mu}_{1(2,2)}}{r^4}\langle J_c^{(1)} \cdot \ell^{(1)}\rangle_\mu + \left(\frac{D^{3\mu}_{2(2,0)}}{r^3} + \frac{D^{4\mu}_{2(2,2)}}{r^4} + \frac{D^{6\mu}_{2(2,4)}}{r^6}\right)\langle J_c^{(2)} \cdot \ell^{(2)}\rangle_\mu$$

$$+ \left(\frac{D^{4\mu}_{3(2,2)}}{r^4} + \frac{D^{6\mu}_{3(2,4)}}{r^6}\right)\langle J_c^{(3)} \cdot \ell^{(3)}\rangle_\mu + \left(\frac{D^{4\mu}_{4(2,2)}}{r^4} + \frac{D^{5\mu}_{4(4,0)}}{r^5} + \frac{D^{6\mu}_{4(2,4)}}{r^6}\right)\langle J_c^{(4)} \cdot \ell^{(4)}\rangle_\mu$$

$$+ \frac{D^{6\mu}_{5(2,4)}}{r^6}\langle J_c^{(5)} \cdot \ell^{(5)}\rangle_\mu + \frac{D^{6\mu}_{6(2,4)}}{r^6}\langle J_c^{(6)} \cdot \ell^{(6)}\rangle_\mu. \tag{4.35}$$

Since the degenerate contributions arise within the context of a single channel adiabatic analysis of Rydberg motion, these terms do not appear in the more general multichannel Hamiltonian treatment of Chapter 3, and are, therefore, not included in the numerical studies presented in Chapters 5 through 7. Here the terms $D^{n\mu}_{k(a,b)}$ corresponding to tensorial rank k, of radial order $\left(\frac{1}{r}\right)^n$, and with electric a-multipole and b-multipole moment contributions are given by

$$D^{n\mu}_{k(a,b)} = \frac{(-1)^{k+2J_c+\ell_\mu}(2k+1)(2\ell_\mu+1)}{\langle \ell_\mu \,||\, \ell^{(k)} \,||\, \ell_\mu\rangle\langle J_c \,||\, J_c^{(k)} \,||\, J_c\rangle}\sum_{\ell_\nu}\Delta^n_{\mu\nu}(a,b)(-1)^{\ell_\nu}(2\ell_\nu+1)$$

$$\begin{pmatrix} \ell_\mu & a & \ell_\nu \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \ell_\nu & b & \ell_\mu \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} a & b & k \\ \ell_\mu & \ell_\mu & \ell_\nu \end{Bmatrix}\begin{Bmatrix} a & b & k \\ J_c & J_c & J_c \end{Bmatrix}$$

$$\langle \gamma_c J_c \,||\, \sum_{i=1}^{N_e} r_i^a C^{(a)}(\hat{r}_i) \,||\, \gamma_c J_c\rangle\langle \gamma_c J_c \,||\, \sum_{j=1}^{N_e} r_j^b C^{(b)}(\hat{r}_j) \,||\, \gamma_c J_c\rangle, \tag{4.36}$$

where the $\Delta^n_{\mu\nu}(a,b)$'s are

$$\Delta^3_{\mu\nu}(2,0) = \Delta^5_{\mu\nu}(4,0) = 1, \tag{4.37}$$

$$\Delta^4_{\mu\nu}(2,2) = \frac{2}{\ell_\mu(\ell_\mu+1) - \ell_\nu(\ell_\nu+1)}, \tag{4.38}$$

$$\Delta^6_{\mu\nu}(2,4) = 2\Delta^4_{\mu\nu}(2,2). \tag{4.39}$$

Like the terms in the nondegenerate case these degenerate coefficients $D^{n\mu}_{k(a,b)}$ exhibit a factorization of ionic core and the Rydberg information. These degenerate coefficients depend on quadrupole and/or hexadecapole moments of the ionic core, but like the vector term, they also depend of the orbital momentum of the Rydberg electron. The additional ℓ_ν dependence generates odd tensorial terms with ranks 3 and 5. However, the combination of quadrupole and hexadecapole moments and the reciprocal ℓ-dependence of the $\Delta^4_{\mu\nu}(2,2)$ term suggests that these degenerate contribution are generally small, but they cannot be neglected if a single channel adiabatic analysis is used to interpret experimental Rydberg energy distributions.

4.5 Tensor Analysis of Diabatic Hamiltonian

Generally, the coupling of nearly degenerate Rydberg states attached to different ionization thresholds of a ionic core results in complex perturbed spectra, which are difficult to treat using a single channel adiabatic formulation. For this reason, the diabatic formulation of Chapter 3 is more robust than the adiabatic approach of Chapter 2 for explicit calculations and for the achievement of spectroscopic accuracy. As with the adiabatic theory we perform a spherical expansion of each term in the diabatic Hamiltonian in powers of $\frac{1}{r}$, and then disentangle all core and Rydberg information through the use of the tensor recoupling techniques presented in section 4.2. However, instead of transforming to an angular momentum operator representation, we retain the unit tensor notation introduced in subsection 4.2.3 so that coupling among the Rydberg-core channels can be properly treated. Furthermore, we treat here the special case of low-Z atomic Rydberg systems. Later in Chapter 7 we present a diabatic formulation for Rydberg states of H_2 and D_2 that assumes no special circumstances and that displays the channel dependence of the core polarizabilities and hyperpolarizabilities.

4.5.1 High ℓ States of Low Z Atoms In low-Z atomic systems where the spin-orbit interaction is small relative to electrostatic interactions, the total orbital angular momentum L_c and the total spin S_c of the core are approximately good quantum numbers. In these systems, the effective diabatic Hamiltonian that describes high-ℓ Rydberg states takes the parametrized form

$$\begin{aligned}\bar{H}^{\mu\mu'} &= \left(-\frac{1}{2}\frac{d^2}{dr^2} + \frac{\ell_\mu(\ell_\mu+1)}{2r^2} - \frac{1}{r} + \Delta E_{\mu\mu_0} - \frac{\alpha_s}{2r^4} - \frac{\eta_s}{r^6}\right)\delta_{\mu\mu'} - \left(\frac{Q}{r^3} + \frac{\alpha_t}{2r^4} + \frac{\eta_t}{r^6}\right)\mathcal{A}^{(2)}_{\mu\mu'} \\ &+ \left(\frac{\beta_s \mathcal{B}^{(0)}_{\mu\mu'} + \beta_t \mathcal{B}^{(2)}_{\mu\mu'}}{r^6}\right) + \frac{\beta_v}{r^6}\langle \vec{L}_c \cdot \vec{\ell}\rangle_{\mu\mu'} \\ &- 2\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{A}^{(2)}_{\mu\mu'}}{r^5}\right) + p_r\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{A}^{(2)}_{\mu\mu'}}{r^4}\right)p_r,\end{aligned} \tag{4.40}$$

where $p_r \equiv -i\frac{d}{dr}$ and where $\Delta E_{\mu\mu_0} \equiv E_\mu - E_{\mu_0}$. Similarly, the matrix $\Lambda_{\mu\mu'}$ that appears on the right-hand side of the generalized eigensystem is given by

$$\Lambda_{\mu\mu'} = \delta_{\mu\mu'} + 2\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{A}^{(2)}_{\mu\mu'}}{r^4}\right). \tag{4.41}$$

(Note: the effective operators $\bar{H}$ and Λ act on the reduced radial wavefunction $\psi_\mu = r\Psi_\mu$ within the P subspace.) The indices μ and μ' refer to channels within the P subspace.

Here we adopt the convention in which α_s and α_t are the standard scalar and tensor dipole polarizabilities, Q is the quadrupole moment, and η_s and η_t are higher-order scalar and tensor hyperpolarizabilities of the ionic core. Explicit expressions for all terms, including the angular factors such as $\mathcal{A}^{(2)}_{\mu\mu'}$ and $\mathcal{B}^{(2)}_{\mu\mu'}$, are given in Appendix A.

The β_s, β_v, and β_t terms that involve $\frac{1}{r^6}$ in Eq.(4.40) represent nonadiabatic scalar, vector, and tensor induced-polarization corrections due to the angular motion of the Rydberg

electron, while the β_s and β_t terms that involve $\frac{1}{r^4}$ and $\frac{1}{r^5}$ represent scalar and tensor nonadiabatic corrections that are associated with the radial motion of the Rydberg electron. The proportionality coefficient β_v (vector hyperpolarizability) of the vector interaction $\vec{L_c} \cdot \vec{\ell}$ deserves special attention. This vector hyperpolarizability depends on the dipole moments of the ionic core and its existence hinges on the centrifugal repulsion experienced by the Rydberg electron. Physically, the vector interaction describes a coupling among the orbital motion of the Rydberg electron and the internal angular momentum of the ionic core, which is mediated by the dipole portion of the Rydberg-core interaction. As the distant Rydberg electron orbits the core, it tries to "drag" the core polarization vector with it. However, this drag is hindered by the internal moment of inertia of the core, and by the moment of inertia of the Rydberg electron. Stated another way, this interaction reflects the inability of the ionic core to instantaneously adjust to the angular motion of the Rydberg electron. The vector nature and the $\frac{1}{r^6}$ behavior of the interaction are independent of representation. However, the core angular momentum operator which appears in $\vec{L_c} \cdot \vec{\ell}$, *does* depend on the choice of representation. Specifically, in an adiabatic representation the core operator is $\vec{J_c}$, while in a diabatic representation the operator is $\vec{L_c}$. This artificial dependence arises because of the difference in the channel coupling appropriate for these representations.

The combination of r-dependence and tensorial structure provide a great deal of insight into the distribution of Rydberg levels. The Coulomb interaction is the dominant term and accounts for the nearly hydrogenic nature of Rydberg systems. Provided the ionic core is nonspherical, the quadrupole interaction, which is the dominant anisotropic term, follows in importance and has the effect of spreading Rydberg levels. Scalar quantities like α_s produce simple shifts of different ℓ-levels, while higher order tensor terms like α_t cause additional spreading, although to a lesser degree than the quadrupole term. The vector term has the effect of further splitting levels that share the same symmetry, but differ in their ℓ dependence. Moreover, the operator $\vec{L_c} \cdot \vec{\ell}$ has the effect of splitting levels that are degenerate

in the quantum number associated with the projection of $\vec{\ell}$ onto $\vec{L_c}$.

The operator structure of the effective Hamiltonian and Λ matrices, and the coupling among P subspace channels that they describe, improves significantly over earlier perturbative formulations. The perturbations among Rydberg levels and the resulting complex spectra can be easily understood and described within this coupled-channel framework. For instance, the wave function of a Rydberg electron may be thought of as distributed among the various channels (paths). The combination of amplitudes from the various channels can result in interference and, consequently, complex spectra. The degree to which this takes place depends on the coupling among channels and whether the channels support strongly overlapping series converging to the various thresholds.

4.6 Qualitative Interpretation of the Vector Interaction

One of the most difficult challenges of our analysis has been to provide a qualitative interpretation of the new vector interaction term in the potential energy. In this section we present a qualitative discussion of this interaction from four perspectives: first by analytically rearranging the operator structure of the adiabatic β_μ^{ad} term, then by considering a two electron example from both a classical and quantum mechanical perspective [31], and finally from the non-Abelian gauge perspective of Zygelman[27].

4.6.1 Analytic Recoupling Analysis In the adiabatic formulation of Chapter 2, only the radial coordinate of the Rydberg electron was treated adiabatically. For this reason, effects arising from Rydberg angular motion can be considered as nonadiabatic. As mentioned in Chapter 2, the vector interaction originated from the radially-adiabatic β_μ^{ad} term. Here we show that this term can be directly recast in an analytic form in which a cross product of the ionic core electric dipole operators appear as a projection onto the orbital momentum of the Rydberg electron. This cross product suggests that torques are present in the dynamics of the nonadiabatic Rydberg-core interaction. In order to see how torques can arise in a Rydberg system, we begin by writing β_μ^{ad} as

$$\beta_\mu^{ad} = \sum_{\nu\neq\mu} \frac{[\ell_\nu(\ell_\nu+1) - \ell_\mu(\ell_\mu+1)]}{(E_\nu - E_\mu)^2} Q^{(1)}_{\mu\nu} Q^{(1)}_{\nu\mu}$$

$$= \sum_{\nu\neq\mu} \frac{\langle\mu \mid \left[\hat{r}_r, \vec{\ell}^2\right] \cdot \vec{r}_c \mid \nu\rangle\langle\nu \mid \hat{r}'_r \cdot \vec{r}'_c \mid \mu\rangle}{(E_\nu - E_\mu)^2}. \tag{4.42}$$

The commutator of $\hat{r}_r$ and $\vec{\ell}^2$ is [51]

$$\left[\hat{r}_r, \vec{\ell}^2\right] = i(\vec{\ell}\times\hat{r}_r - \hat{r}_r\times\vec{\ell}), \tag{4.43}$$

and with a little tensor recoupling the term with a vector structure in β_μ^{ad} is seen to be proportional to

$$\left(\left[\hat{r}_r, \vec{\ell}^2\right] \times P_r\hat{r}'_r\right) \cdot (\vec{r}_c \times P_c\vec{r}'_c) \tag{4.44}$$

Using

$$\left[\hat{r}_r, \vec{\ell}^2\right] \times P_r\hat{r}'_r = -2i(\hat{r}_r \cdot P_r\hat{r}'_r)\vec{\ell} + 2i\hat{r}_r(\vec{\ell}\cdot P_r\hat{r}'_r) - 2\hat{r}_r \times P_r\hat{r}'_r \tag{4.45}$$

we see the appearance of a $\frac{1}{r^3}$ vector interaction that is proportional to

$$\sim \vec{\ell}\cdot(\vec{r}_c \times P_c\vec{r}'_c) \tag{4.46}$$

The existence of this vector interaction hinges on the presence of the centrifugal $\ell_\nu(\ell_\nu+1)$ term within β_μ^{ad}, which again supports our conclusion that the vector interaction results from the angular motion of the Rydberg electron. In addition, the vector cross product on the right hand side of Eq.(4.43) suggests that the Rydberg electron exerts a torque on the ionic

core as the Rydberg electron attempts to "drag" the polarization vector of the spinning core. This idea of the Rydberg electron torquing the core will be more fully explored in chapter 6 where the concept of orbital planes is introduced, and the relative orientation of the core and Rydberg orbital planes is studied within the adiabatic representation [57, 58].

4.6.2 A Two-Electron Example Treated Classically Next we turn to a classical example with two nonoverlapping coplanar electrons that move in the field of a $Z = 2$ charged nucleus as shown in Figure 4.1. The inner electron is initially in a circular Bohr orbit with $n_1 = 2$ and $\ell_1 = 1$, while the outer electron is started in a circular orbit with $n_2 = 7$ and $\ell_2 = 6$. The zeroth-order Hamiltonian for this system then looks like:

$$H_0 = \frac{1}{2}\vec{p_1}^2 + \frac{1}{2}\vec{p_2}^2 - \frac{2}{r_1} - \frac{1}{r_2}. \tag{4.47}$$

We assume here that $r_2 >> r_1$, so that the inner electron ($r_1 \approx 2$) fully screens the outer electron ($r_2 \approx 49$), and the leading order perturbation to H_0 is the dipole component of the electron-electron interaction:

$$V = \frac{r_1 \cos\theta_{12}}{{r_2}^2}. \tag{4.48}$$

In order to find the energy shift of the system due to the perturbative dipole interaction V, we propagate classical trajectories for the Hamiltonian $H_0 + R(t)V$, where R(t) is an adiabatic "ramp-on" that is zero at time $t = 0$ and becomes unity after many orbit periods of the outer electron. (This method of evaluating perturbation energies in conjunction with classical mechanics has been used successfully in a different context by Hooker [59].) Figure 4.2 shows that the resulting perturbation in the energy of the system has opposite sign for the two geometries of Figure 4.1. In particular, the energy shift is negative when the two angular momenta $\vec{\ell_1}$ and $\vec{\ell_2}$ are parallel (left-hand case in Figure 4.1, while the energy shift has nearly the same magnitude but is positive when these angular

momenta are antiparallel (right-hand case in Figure 4.1).

4.6.3 Two-Electron Example Treated Using Quantum Mechanics The same system just considered classically possesses a simple quantum mechanical description. One quantitative difference between this example of a one-electron ionic core in the $n_1 = 2$ state of He^+ and the cases formulated in Chapters 2 and 3, is that a one-electron hydrogenic ion possesses the unusual "accidental degeneracy" (when fine structure is neglected) of the $2s$ and $2p$ substates. It is well known that this degeneracy magnifies the effect of the dipole interaction between a distant electron and the hydrogenic ion, which is one reason we have chosen this example. At the same time, this system can be treated simply using 2x2 matrices in an LS-coupled representation. In the following, we let r stand for r_2, the distance of the outer electron from the nucleus, while ℓ is the orbital angular momentum of that outer electron. The leading order term at $r \to \infty$ in the potential accordingly contributes in order r^{-2}, *i.e.* on a par with the centrifugal term in the Rydberg electron Hamiltonian.

The quantum mechanical channel describing this $2p n\ell$ two electron atom thus has the following character. When the angular momenta of the inner ($\ell_1 = 1$) and outer electron ($\ell = 6$) are "parallel", the zeroth-order state can be written as $|(2p, n\ \ell = 6)L = 7\rangle$, i.e. as an eigenstate of $\vec{L}^2 = L(L+1)$ with $L = 7$. The dominant channel that couples to this state is the channel $|(2s, n\ \ell + 1 = 7)L = 7\rangle$. The correction to the long-range Coulomb potential of the outermost electron is therefore a dipolar $\propto \frac{1}{r^2}$ potential for this example, whose coefficient is obtained by diagonalization of the combined centrifugal and dipole-interaction potentials. Through dipolar order, neglecting terms of radial order $\frac{1}{r^3}$, the effective potential appropriate for parallel angular momenta is the smaller eigenvalue of:

$$V^{eff}(r) = -\frac{1}{r} + \begin{pmatrix} \frac{\ell(\ell+1)}{2r^2} & \frac{d}{r^2} \\ \frac{d}{r^2} & \frac{(\ell+1)(\ell+2)}{2r^2} \end{pmatrix}.$$

Here d is a matrix element of the core electric dipole operator between the two coupled channels, $d = \langle 2p|r_1|2s\rangle\langle(1\ell)L = \ell + 1|\cos\theta_{12}|(0, \ell + 1)L = \ell + 1\rangle = -\sqrt{\frac{21}{20}}$. The analogous

effective potential relevant for antiparallel angular momenta of the core and the Rydberg electron is the larger eigenvalue of:

$$V^{eff}(r) = -\frac{1}{r} + \begin{pmatrix} \frac{\ell(\ell+1)}{2r^2} & \frac{d'}{r^2} \\ \frac{d'}{r^2} & \frac{(\ell-1)\ell}{2r^2} \end{pmatrix},$$

where $d' = \langle 2p|r_1|2s\rangle\langle(1\ell)L = \ell - 1|\cos\theta_{12}|(0,\ell-1)L = \ell - 1\rangle = \sqrt{\frac{27}{22}}$. The resulting eigenvalues give effective potentials for the outer ($\ell = 6$) electron:

$$V^{eff}_{\uparrow\uparrow}(r) = -\frac{1}{r} + \frac{\ell(\ell+1)}{2r^2} - \frac{0.147}{r^2},$$

$$V^{eff}_{\uparrow\downarrow}(r) = -\frac{1}{r} + \frac{\ell(\ell+1)}{2r^2} + \frac{0.198}{r^2}.$$

The last term in each potential gives the effect of the perturbation. The quantum mechanical estimate of the perturbation energy is obtained by averaging the last term in the above two potentials over the radial $n = 7, \ell = 6$ orbital, which gives:$\Delta^{QM}_{\uparrow\uparrow} = -6.6\times10^{-5} a.u.$, and $\Delta^{QM}_{\uparrow\downarrow} = 8.9 \times 10^{-5} a.u.$ These quantum mechanical energy shifts for this example are within about a factor of two of the classically-estimated energy shifts, confirming that the basic classical picture presented above is qualitatively correct. Interestingly, the quantum mechanical interpretation for this prototype two-electron example shows that the different sign of the energy perturbation for parallel and antiparallel angular momenta is a consequence of differential level repulsion. The state $2pn\ell[L = \ell + 1]$, with parallel angular momenta, is pushed down in energy by its interaction with the higher-lying $2sn'(\ell+1)[L = \ell + 1]$ state, whereas the state $2pn\ell[L = \ell - 1]$, with antiparallel angular momenta, is pushed up in energy by the lower $2sn'(\ell-1)[L = \ell - 1]$ state. This qualitative interpretation remains similar when the ionic s and p levels are not degenerate, except that the differential level repulsion then causes the ℓ-dependence in the potential to begin at the r^{-6}level rather than the r^{-2} level.

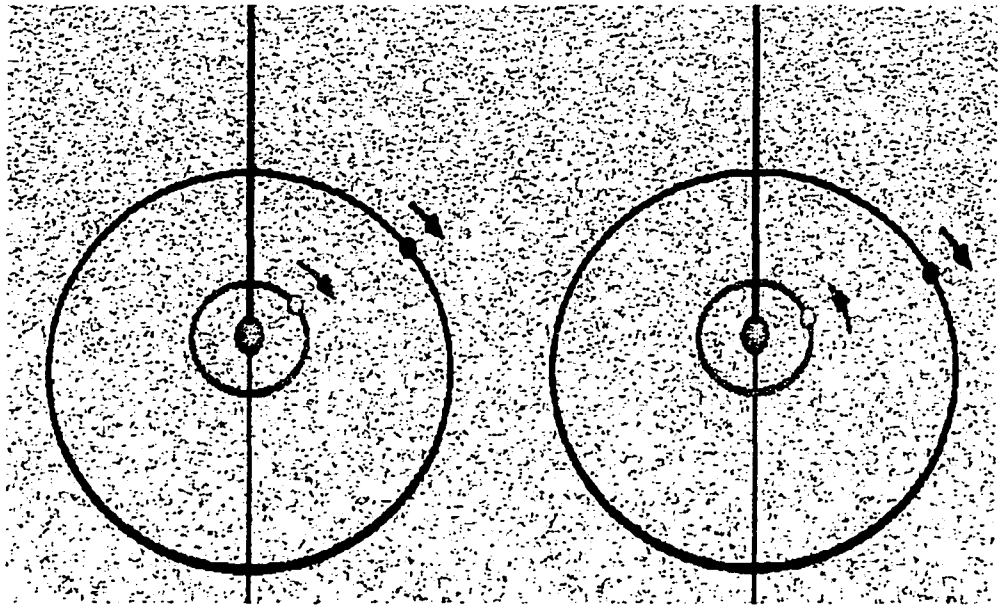


Figure 4.1. Classical orbits of a Rydberg electron around a core consisting of an electron with angular momentum circling about a nucleus. The interaction energy associated with the vector interaction term has opposite sign depending on whether the angular momenta are parallel (left) or antiparallel (right).

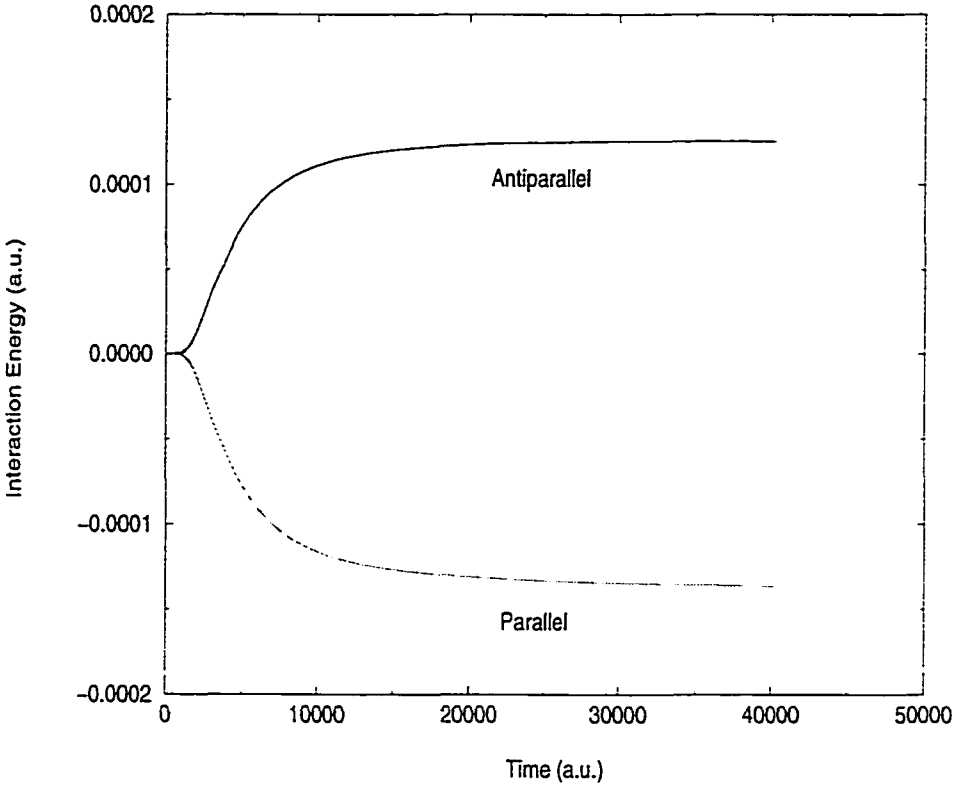


Figure 4.2. Resulting perturbation in the energies of the parallel and antiparallel classical orbits versus time.

4.7 Non-Abelian Gauge Formulation

Zygelman [27] was the first to predict the existence of the vector interaction and to estimate the magnitude of the proportionality constant. In an original and pioneering approach, Zygelman used ideas of non-Abelian gauge fields and geometric phases to study long range atomic forces. Extending the work of Berry [60], Wilczek and Zee[61], Moody [62] and Jackiw [63] Zygelman realized that the derivative coupling matrix $\underline{P}(r)$ within an adiabatic representation resembles an electromagnetic vector potential or a non-Abelian gauge field,

$$\left[-\frac{1}{2}\left(\underline{I}\frac{d}{dr}+\underline{P}(r)\right)^2-(E\underline{I}-\underline{U}(r))\right]\underline{F}(r)=0.$$

As in the classical dynamics of charged particles in electromagnetic fields, [27, 37, 38], a gauge transformation can in principle be found that gauges away the derivative coupling matrix. However, if the channel space is finite or truncated the derivative coupling can be only approximately gauged away, leaving, as in our formulation, an effective potential. It was within this framework that Zygelman first suggested that a vector interaction involving $\vec{L_c}\cdot\vec{\ell}$ does exist as a long range interaction between a charge and an anisotropic system. However, an explicit expression for the coefficient of proportionality was not provided, leaving questions regarding the physical origin of this interaction and to some extent even the fundamental reality of this term, since the coefficient could be zero.

CHAPTER 5

RYDBERG STATES OF NEON

In this Chapter we present the first experimental and theoretical confirmation of the vector interaction in a Rydberg system. The analysis presented here both establishes the importance of this subtle interaction and demonstrates the utility of an effective Hamiltonian approach.

5.1 Application to Neon

The importance of the vector interaction and the channel coupling provided by the effective Hamiltonian approach have been demonstrated in the $n = 10$ Rydberg states of Ne with orbital angular momenta of $\ell = 5, 6, 7$ and 8. These states have been studied in recent experiments by Ward *et al* [10], and interpreted by our theoretical analysis[30]. The most important aspect of this analysis is the capability of our parametrized long-range potential to describe the Rydberg energy levels that are attached to the lowest ionization threshold of the Ne^+ ionic core. The theoretical computation of Rydberg energies is accomplished by diagonalization of the multichannel Hamiltonian

$$H_{\mu\mu'} = \left(-\frac{1}{2}\frac{d^2}{dr^2} + \frac{\ell_\mu(\ell_\mu+1)}{2r^2} - \frac{1}{r} + E_\mu - \frac{\alpha_s}{2r^4} - \frac{\eta_s}{r^6}\right)\delta_{\mu\mu'}$$

$$- \left(\frac{Q}{r^3} + \frac{\alpha_t}{2r^4} + \frac{\eta_t}{r^6}\right)\mathcal{A}^{(2)}_{\mu\mu'} + \frac{\beta_v}{r^6}\langle\vec{L}_c\cdot\vec{\ell}\rangle_{\mu\mu'}, \tag{5.1}$$

followed by a minimization of the weighted χ^2 function (involving differences between the observed and computed energy levels) with respect to the parameters α_s, α_t, β_v, $\cdots$. This version of the effective Hamiltonian can be easily derived from our more general diabatic

Hamiltonian by treating ε, in the Green's functions expansion, as a hydrogenic Hamiltonian. We do this here so that the diagonal elements of our multichannel Hamiltonian exactly correspond with the perturbative Hamiltonian used in the Ward analysis [10]. It is worth noting that there is nothing *ad-hoc* in the parametrized theory. Analytic expressions for each parameter are given in Refs.[30, 31] and in Appendix A, and each can be computed from first principles. In fact, *ab initio* calculations of α_s, α_l, β_v, $\cdots$ using multiconfiguration Hartree-Fock atomic wavefunctions, presented in the next section, confirm the analysis and the interpretation presented here.

The experimentally observed spin splittings were resolved only partially in the Ward *et al* experiment and were analyzed by the experimental group to extract spinless transition frequencies. We neglect all spin-orbit terms in the Rydberg electron Hamiltonian, and all Rydberg electron exchange interactions. The dominant perturbative effects that must be added to our computed Rydberg levels are the relativistic "mass correction" arising from the p^4 term in the kinetic energy of the Rydberg electron, and the magnetic interaction between the Rydberg electron and the ionic core given by

$$H_{mag} = -\frac{g_J}{2}\alpha^2 \frac{\vec{J}_c \cdot \vec{\ell}}{r^3}, \tag{5.2}$$

where α is the fine-structure constant and g_J is the g factor of the ionic core.

The results of our analysis are summarized by Tables 5.1 and 5.2. Table 5.1 compares the various fitted parameters for Ne^+ with those obtained in a fit carried out by Ward *et al* [10]. Our analysis, which nonperturbatively includes coupling between Rydberg series attached to different ionization thresholds, improves the χ^2 by nearly 80% over the fit performed by Ward *et al.*. The fact that the χ^2 per degree of freedom (per parameter) is now approximately *one* is a strong indication that the anisotropic multichannel Hamiltonian correctly accounts for the energetics of these Rydberg levels. Moreover, it also suggests that our nonperturbative treatment of channel coupling is an improvement on the perturbative

description of Rydberg levels used by Ward *et al*, since our fit and their fit were based on the same Hamiltonian, apart from the vector contribution to the potential energy. Our computed levels are compared in Table 5.2 with those of Ref. [10], for the n = 10 Rydberg states of Ne with $\ell = 5, 6, 7$ and 8. These parameters reproduce all observed energy levels to 0.5 MHz or better.

In order to determine the importance of the vector interaction, it is instructive to repeat the same least-squares fitting analysis described in the previous paragraph, but with the constraint that the vector hyperpolarizability β_v must vanish. Table 5.1 lists the various fitted parameters obtained from this analysis without the vector interaction. The greatest change among the parameters occurs in the Ne^+ $^2P_{\frac{3}{2}}$ gyromagnetic ratio g_J, which changes from 1.342 to 1.307. The tensorial structure of the gyromagnetic term is similar to that of the vector interaction and it is plausible that the resulting fit modifies g_J to "mock up" the effects omitted by setting β_v equal to zero. However, since both the radial dependence of these two interactions are different, a change in g_J can only approximately compensate for the omission of β_v from our nonlinear fit. Note that the LS coupling value of g_J is precisely $\frac{4}{3}$, but the effects spin-orbit coupling can change this value from this limiting LS-coupled value. (An independent, linear Zeeman effect measurement could easily test the fitted g_J value.)

The Rydberg levels computed without the vector interaction are given in Table 5.2. The largest discrepancies among the computed and observed levels appear in the lower ℓ states, where the largest deviation is -1.43 MHz in $(\frac{3}{2})10H_{\frac{9}{2}}$. A clear difference is apparent, though, between this analysis and that obtained with the vector hyperpolarizability included. The new least-squares fit obtained without the vector interaction results in $\chi^2 = 27.5$, a χ^2 per degree of freedom approximately four times worse than the χ^2 for the fit including the vector interaction. We interpret this as strong evidence that the existence of the vector interaction has been confirmed in the Ward *et al* experiment on Ne Rydberg states.

Table 5.1. Comparison of fitted (experimental) and theoretical parameters for Ne^+. L refers to length form and V refers to velocity form. The value of g_J in column 4 is based on pure LS-coupling*. The values labelled "present" are from our earlier work (Clark *et al.* 1996), as are the unlabelled theoretical values for α_s, α_t, and β_v.

	Present Fit	Present Fit without β_v	Ward *et.al.* fit	Theoretical
Q	-0.204020(5)	-0.204001(11)	-0.20403(5)	-0.1964 [66] -0.2032(5) [66] -0.2117 [10]
α_s	1.3018(2)	1.3011(6)	1.3028(13)	1.23 (L) 1.19 (V) 1.27 [65]
α_t	-0.0259(3)	-0.0261(3)	-0.026(5)	-0.0374 (L) -0.0396 (V) -0.035 [10]
β_v	0.059(2)	'0'	0.045(29)	0.0678 (L) 0.0719 (V)
η_s	-0.10(1)	-0.10(1)	-0.29(24)	-1.44 [67]
η_t	0.274(5)	0.264(3)	0.5(5)	
g_J	1.342(12)	1.307(24)	1.354(21)	$\frac{4}{3}$ *
χ^2	7.1	27.5	35.7	

Table 5.2. Comparison of calculated (with and without β_v, see Clark *et al.* (1996)) and experimentally observed (Ward *et al.* 1996) energies (MHz) of $n = 10$ Rydberg neon with $J_c = \frac{3}{2}$ and $\ell = 5, 6, 7$, and 8. $\Delta E = E_{obs} - E_{calc}$.

States	$E_{obs}(MHz)$ [7]	E_{calc}	ΔE	$E_{calc}^{\beta_v=0}$	$\Delta E^{\beta_v=0}$
$H_{\frac{9}{2}}$	-145.58(77)	-145.63	0.05	-144.15	-1.43
$H_{\frac{11}{2}}$	2142.67(10)	2142.60	0.07	2142.81	-0.14
$H_{\frac{13}{2}}$	-6022.24(19)	-6022.02	0.22	-6022.66	0.42
$I_{\frac{9}{2}}$	-5267.15(35)	-5267.38	0.23	-5266.64	-0.51
$I_{\frac{11}{2}}$	-356.30(24)	-356.18	-0.12	-355.91	-0.39
$I_{\frac{13}{2}}$	800.52(5)	800.50	0.02	800.55	-0.03
$I_{\frac{15}{2}}$	-4131.36(15)	-4131.35	-0.01	-4131.10	-0.26
$K_{\frac{11}{2}}$	-3838.06(35)	-3838.50	0.44	-3838.26	0.20
$K_{\frac{13}{2}}$	-646.41(8)	-646.37	-0.04	-646.36	-0.05
$K_{\frac{15}{2}}$	'0'	'0'	'0'	'0'	'0'
$K_{\frac{17}{2}}$	-3205.01(16)	-3204.97	-0.04	-3204.68	-0.33
$L_{\frac{13}{2}}$	-3073.14(35)	-3073.64	0.50	-3073.56	0.42
$L_{\frac{15}{2}}$	-883.09(8)	-883.04	-0.05	-883.08	-0.01
$L_{\frac{17}{2}}$	-494.04(5)	-494.04	0.00	-494.05	0.01
$L_{\frac{19}{2}}$	-2693.41(18)	-2693.38	-0.03	-2693.14	-0.27

To complement our analysis of this Rydberg system we present *ab initio* theoretical values for α_s, α_t, and β_v, in Table 5.1 for comparison with other theoretical and experimental results. Our theoretical results are presented in both the length and velocity forms of the electric dipole matrix elements. The details of these calculations are presented in the next section. In view of the difficulty of the *ab initio* calculation of such parameters for a many-electron ion, the calculated values for Ne^+ are in reasonably good agreement, both in magnitude and in sign, with the values extracted from the measurements of Ward *et al* [10]. This agreement strengthens the argument that the vector interaction does exist and has experimental implications in certain features of Rydberg spectra.

As another demonstration of the importance of channel coupling in the Rydberg states of Ne, we display in Figures 5.1 and 5.2 two sets of adiabatic potentials for Ne, which exhibit clear qualitative differences. In Figure 5.1 the adiabatic potentials correspond to $K^\pi = \frac{9}{2}^-$ and are labeled from top to bottom with $(J_c, \ell) = (\frac{1}{2}, 5)$, $(\frac{3}{2}, 5)$, and $(\frac{3}{2}, 3)$. The absence of avoided crossings and the smoothly decaying behavior of the derivative couplings suggest that an adiabatic analysis is justifiable in this case. However, Figure 5.2 demonstrates that adiabatic potentials for different symmetries can differ in a crucial manner. In Figure 5.2 the adiabatic potentials corresponding to $K^\pi = \frac{11}{2}^-$ and are labeled from top to bottom with $(J_c, \ell) = (\frac{1}{2}, 5)$, $(\frac{3}{2}, 7)$, and $(\frac{3}{2}, 5)$. In this case, the presence of a close avoided crossing and a corresponding sharp derivative coupling indicate that a simple adiabatic analysis cannot account for the strong coupling among the nearly degenerate channels. For these reasons, the effective diabatic Hamiltonian approach is preferable, as has been demonstrated in this section.

It should be noted, though, that an adiabatic analysis can be very useful for developing a qualitative understanding of simple and complex spectra. For example, the wave function of a Rydberg electron moving in the presence of these potentials will be distributed among the various channels (paths). The coherent superposition of amplitudes from the

various paths can lead to interference that results in complex spectra. The extent to which such perturbations influence the spectrum depends on the coupling between channels, and on whether the Rydberg series converging to the various ionization thresholds overlap extensively.

5.2 First Calculation of the Vector Hyperpolarizability β_v

In this section we discuss the calculation of the reduced dipole matrix elements necessary to evaluate the vector hyperpolarizability β_v, the standard dipole scalar polarizability α_s, and the second-rank tensor polarizability α_t of Ne^+. The theoretical values for α_s, α_t, and β_v are presented in both length and velocity form in Table 5.1 for comparison with other theoretical and experimental results. Reduced dipole matrix elements are calculated for the ground state ($2s^2 2p^5\ ^2P^o$) of Ne^+. The dipole operator $r^{(1)}C_q^{(1)}$ connects states of the opposite parity which differ by at most one orbital, and such that $\Delta L_c = 0, \pm 1$ (except for $L_c = 0$ to $L_c = 0$ transitions) and $\Delta S_c = 0$. Thus only the ^2S, ^2P, and ^2D final states are needed, which are generated from the ground state by $2s \rightarrow np$, $2p \rightarrow ns$ and $2p \rightarrow nd$ substitutions. Excitations of the $1s$ core are ignored since these give negligible contributions to reduced dipole matrix elements. These final states of Ne^+ are represented by $2s2p^6$, $2s2p^5np$, $2s^2 2p^4 ns$ and $2s^2 2p^4 nd$ configurations, which can be constructed (including the ground state $2s^2 2p^5$) from a product of Ne^{2+} states $2s2p^5$ or $2s^2 2p^4$, and an outer s, p or d electron. These Ne^{2+} configurations are referred to as physical target states.

The summations over bound and continuum states of Ne^+ are accomplished using the eigenchannel R-matrix method [55]. This allows us to construct a complete set of orthogonal basis functions, vanishing at the boundary of the R-matrix sphere of radius r_0. Such functions represent a bound spectrum and a discretized continuum of Ne^+. The target states of Ne^{2+} are calculated using the multiconfiguration Hartree-Fock approximation [68], in which both spectroscopic and correlation orbitals are included. First, spectroscopic

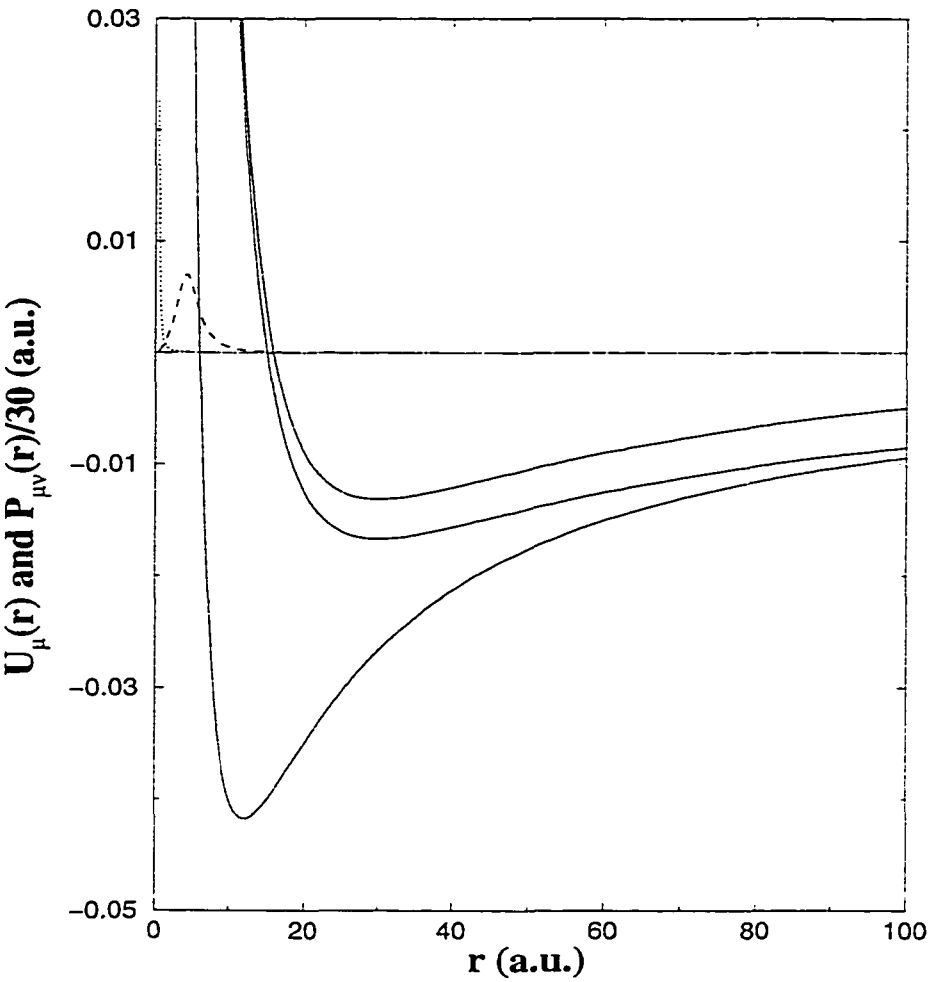


Figure 5.1. Adiabatic potential curves and derivative couplings for Rydberg Ne. The adiabatic potentials correspond to $K^\pi = \frac{9}{2}^-$ and are labeled from top to bottom with $(J_c, \ell) = (\frac{1}{2}, 5)$, $(\frac{3}{2}, 5)$, and $(\frac{3}{2}, 3)$. The derivative couplings $P_{\mu\mu'}(r)/30$ are given by the broken lines: $P_{top,middle}(r)$, dashed; $P_{middle,bottom}(r)$, dotted; and $P_{top,bottom}(r)$, dot-dashed

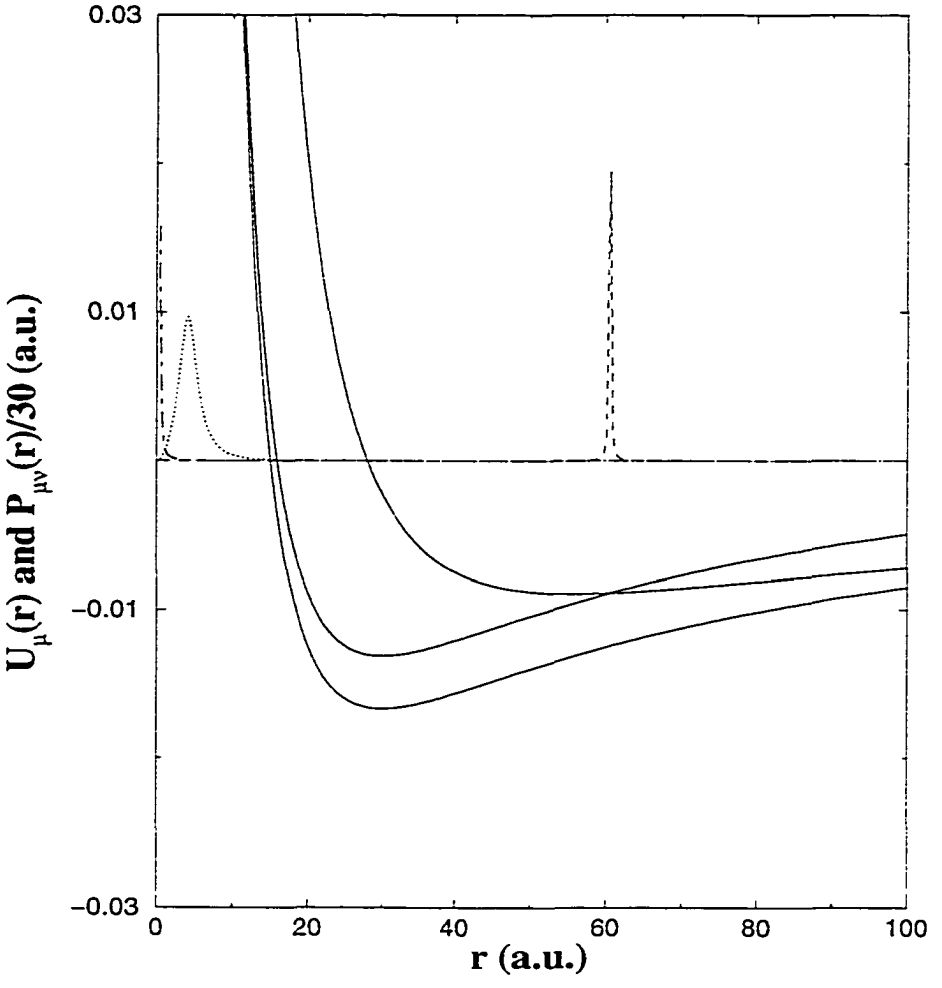


Figure 5.2. Adiabatic potential curves and derivative couplings for Rydberg Ne. The adiabatic potentials correspond to $K^\pi = \frac{11}{2}^-$ and are labeled from top to bottom with $(J_c, \ell) = (\frac{1}{2}, 5)$, $(\frac{3}{2}, 7)$, and $(\frac{3}{2}, 5)$. The derivative couplings $P_{\mu\mu'}(r)/30$ are given by the broken lines: $P_{top,middle}(r)$, dashed; $P_{middle,bottom}(r)$, dotted; and $P_{top,bottom}(r)$, dot-dashed.

orbitals $1s$, $2s$ and $2p$ are optimized on a single $2s^22p^4$ configuration. Then, a correlation $\underline{3d}$ orbital is optimized on $2s2p^5\,{}^3P^o$, whose CI expansion includes the main perturber $2s^22p^33d\,{}^3P^o$. Finally, $\underline{3s}$ and $\underline{3p}$ correlation orbitals are optimized on $2s^22p^4\,{}^3P$, where singly- and doubly-excited configurations allowed by parity and spin-angular momentum coupling rules are included.

We then set up an initial CI expansion for each LS-term in the $2s^22p^4$ and $2s2p^5$ configuration, including all allowed (by parity and LS-symmetry conservation rules) singly- and doubly-excited configurations of $2s$, $2p$, $\underline{3s}$, $\underline{3p}$ and $\underline{3d}$ orbitals. After diagonalizing a Hamiltonian, this initial CI set is condensed; from each eigenvector representing a physical target state, we delete those configurations whose weight is less than 0.0005. In the final step, extra configurations representing the target polarization are added. These are constructed from single-electron excitations from $2s^22p^4$ and $2s2p^5$, involving a change in the parity as $2s \rightarrow 2p$, $2p \rightarrow 3s$ and $2p \rightarrow 3d$. Table 5.3 shows energies and dominant configurations for each physical target state in the condensed basis. Comparing with experiment [69], our relative energies are accurate to at least 5%. As in previous eigenchannel R-matrix calculations [55], a discretized basis of outer-electron orbitals ns, np, nd, nf and ng is used. These are determined inside the R-matrix sphere of radius $r_0 = 7$ Bohr radii. The size of r_0 is chosen to contain all physical target states and the ground state of Ne^+, and to ensure an exponential decay of $rP_{nl}(r)$, where $P_{nl}(r)$ are the ground-state radial functions. A set of radial basis functions for the outermost electron is obtained from a Hartree equation

$$\left(-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} - \frac{Z}{r} + V_H(r)\right) P_{nl}(r) = E_{nl}P_{nl}(r) + \sum_{n'} \lambda_{nn'} P_{n'l}(r),$$

where

$$V_H(r) = \sum_{n'} q_{n'l} \left[\int_0^r \frac{1}{r} P_{n'l}^2(s)ds + \int_r^\infty \frac{r^2}{s^2} P_{n'l}^2(s)ds\right].$$

Here, $q_{n'l}$ are occupation numbers of spectroscopic orbitals representing the $2s^22p^4$ target, and $\lambda_{nn'}$ are Lagrange multipliers needed to orthogonalize the outer-electron and target orbitals (including the correlation orbitals).

All new "box" orbitals are forced to vanish at the R-matrix surface. These constitute a complete orthogonal basis, representing an electron outside the residual Ne^{2+} ion. Those orbitals, which have positive energies $E_{nl} > -\frac{Z_{eff}}{r_0} + \frac{l(l+1)}{2r_0^2}$ not only describe bound states but also represent a discretized continuum of the Ne^+ spectrum. The ground state of Ne^+ is constructed from an antisymmetrized product of the target states and outer-electron orbitals. The energies and atomic wave functions of Ne^+ are just eigenvalues and eigenvectors, respectively, of the Hamiltonian. The ionization energy of $Ne^+(2s^22p^5\,^2P^o)$ obtained in this calculation is 334460 cm^{-1}, whereas the experimental value is 331350 cm^{-1}. The accuracy can be tested more thoroughly by analyzing errors in the whole calculated Rydberg series, instead of a singel level. The theoretical and experimental effective quantum numbers of a $2p$ electron in $2s^22p^5\,^2P^o$ are 0.573 and 0.572, respectively, giving a difference of 0.001 in the quantum defect.

The final ^2S, ^2P and ^2D states are constructed in the same way as the ground state. However, a similar estimate of errors can only be made for the lowest eigenstates whose atomic wave functions fit inside 7 Bohr radii and thus represent physical states of Ne^+. For the lowest, even-parity state $2s2p^6\,^2S$ we obtain an excitation energy of 215953 cm^{-1}, whereas the experimental energy is 217050 cm^{-1}. The corresponding error in the quantum defect is only 0.01, well within the range of errors expected for such a strongly correlated state. Note that $2s2p^6\,^2S$ is correlated predominantly with a $2s^22p^4\underline{3d}\,^2S$ perturber, which contributes nearly 25% of the CI expansion. No similar error analysis can be carried out for higher excited states, of course, since those no longer fit within the R-matrix box. Our final results are obtained with 11 orbitals for each angular momentum l. However, a different number of the box orbitals was initially used to test convergence of the dipole scalar polarizability. For

Table 5.3. Theoretical and experimental energies Ref.[24] in cm^{-1} (upper and lower entry in first column, respectively) of some Ne^{2+} states, relative to the ground state $2s^2 2p^4\ ^3P$, and shortened CI expansions for each of these states.

Energy	Composition		
0.0	$2s^2 2p^4\ ^3P$	$2s2p^4(\ ^2D)3d\ ^3P$	$2s2p^4(\ ^2P)3d\ ^3P$
0.0	0.98605	0.00410	0.00160
25559	$2s^2 2p^4\ ^1D$	$2s2p^4(\ ^2P)3d\ ^1D$	$2s^2 2p^2(\ ^1D)3p^2(\ ^3P)\ ^1D$
25521	0.98406	0.00672	0.00176
53096	$2s^2 2p^4\ ^1S$	$2p^6\ ^1S$	$2s^2 2p^2(\ ^1S)3d^2(\ ^1S)\ ^1S$
55427	0.95453	0.03386	0.00303
203471	$2s2p^5\ ^3P$	$2s^2 2p^3(\ ^2D)3d\ ^3P$	$2s^2 2p^3(\ ^2P)3d\ ^3P$
204589	0.97417	0.01000	0.00397
291435	$2s2p^5\ ^1P$	$2s^2 2p^3(\ ^2D)3d\ ^1P$	$2s^2 2p^3(\ ^2P)3d\ ^1P$
289159	0.96629	0.01300	0.00314

the ^2S and ^2P symmetries, whose calculations consume the least CPU-time and memory, we increased the number of box states to 13. This changed the scalar polarizability by about 2%. We also tested the importance of g-waves, since these, along with f-waves, were neglected in earlier theoretical calculations [65]. The g-waves were found to contribute about 3% to the scalar dipole polarizability, and even greater effects are expected from f-waves. Therefore, these are kept in the present calculations.

Our final value of the scalar polarizability is 1.23 a.u. in length form and 1.19 a.u. in velocity form, which should agree if the wavefunctions are exact eigenfunctions of the Hamiltonian. The scalar polarizability in length form differs by about 6% from the experimental results and by about 3% with other theoretical results (see Table 5.1). In velocity form, the scalar polarizability differs by about 9% with experimental results and by about 6% with theoretical results. The slightly better results obtained in Ref.[65] can be attributed to the implicit summation over Rydberg series and continuua, whereas our method uses a CI expansion which in general converges more slowly.

CHAPTER 6

AUTOIONIZING RYDBERG STATES OF MAGNESIUM

In this chapter we continue our investigation of anisotropic interactions by considering the doubly-excited, autoionizing $3pnf$ states of Mg. Here a Rydberg electron interacts with an excited Mg^+ $3p$ ionic core. In this system, strong coupling among Rydberg states results in irregular behavior in the autoionization rates; we show that the irregularity is accurately described by the effective diabatic Hamiltonian formulation. As discussed in Chapter 4 we explore further the idea that the electric dipole moment of the ionic core experiences a torque as the distant Rydberg electron "revolves" around the core.

6.1 Modification for Low ℓ Rydberg States

In most circumstances, an accurate description of low-ℓ Rydberg states is difficult to achieve using a long-range adiabatic potential or an effective Hamiltonian. The difficulty arises because the short-range physics is described inadequately. However, there are special symmetries of Rydberg systems that are amenable to the methods we have developed. The $3p_{\frac{1}{2}}nf_{\frac{7}{2}}(J=4)$, $3p_{\frac{3}{2}}nf_{\frac{7}{2}}(J=4)$, and $3p_{\frac{3}{2}}nf_{\frac{9}{2}}(J=4)$ Rydberg states of Mg, with $n \geq 9$, fall into this latter category [32]. For these symmetries an f Rydberg electron can only couple to other f and h states, which means that the Rydberg-core interaction can still be treated perturbatively.

For f and g Rydberg states the inner classical turning point ranges from $6-10a.u.$. Exchange can still be reasonably neglected, but the asymptotic expansion of the Rydberg-core interaction in powers of $\frac{1}{r}$ becomes problematic when r becomes comparable to the physical size of the ionic core. In order to treat this small r region better it is preferable to expand in $r_< = min\{r, r_i\}$ and $r_> = max\{r, r_i\}$, where r_i is the radial position of a core

electron.

An analysis identical to that given in Chapters 3 and 4 for the diabatic Hamiltonian can be carried out for this case, with the simple modification that we perform a spherical expansion in $r_<$ and $r_>$. This expansion necessarily makes all electric multipole moments and induced polarizabilities radially-dependent, but this dependence rapidly falls off for $r > 10a.u.$. The effective diabatic Hamiltonian for low-ℓ states of low-Z atoms is

$$\begin{aligned}
\bar{H}^{\mu\mu'} &= \left(-\frac{1}{2}\frac{d^2}{dr^2} + \frac{\ell_\mu(\ell_\mu+1)}{2r^2} + \frac{C_c(r)}{r} + \Delta E_{\mu\mu_0} - \frac{\alpha_s(r)}{2r^4} - \frac{\eta_s(r)}{r^6}\right)\delta_{\mu\mu'} \\
&- \left(\frac{Q(r)}{r^3} + \frac{\alpha_t(r)}{2r^4} + \frac{\eta_t(r)}{r^6}\right)\mathcal{A}^{(2)}_{\mu\mu'} \\
&+ \left(\frac{\beta_s(r)\mathcal{B}^{(0)}_{\mu\mu'} + \beta_t(r)\mathcal{B}^{(2)}_{\mu\mu'}}{r^6}\right) + \frac{\beta_v(r)}{r^6}\langle\vec{L}_c\cdot\vec{\ell}\rangle_{\mu\mu'} \\
&- 2\left(\frac{\beta_s(r)\delta_{\mu\mu'} + \beta_t(r)\mathcal{A}^{(2)}_{\mu\mu'}}{r^5}\right) + p_r\left(\frac{\beta_s(r)\delta_{\mu\mu'} + \beta_t(r)\mathcal{A}^{(2)}_{\mu\mu'}}{r^4}\right)p_r, \qquad (6.1)
\end{aligned}$$

where $\Delta E_{\mu\mu_0} \equiv E_\mu - E_{\mu_0}$. Similarly, the matrix $\Lambda_{\mu\mu'}$ is

$$\Lambda_{\mu\mu'} = \delta_{\mu\mu'} + 2\left(\frac{\beta_s(r)\delta_{\mu\mu'} + \beta_t(r)\mathcal{A}^{(2)}_{\mu\mu'}}{r^4}\right). \qquad (6.2)$$

The additional term $C_c(r)$ represents the partially screened Coulomb charge seen by the Rydberg electron as a function of the radial coordinate. These channel independent radial parameters converge to *zero* in the limit $r \to 0$, except for $C_c(r)$ which converges to $-Z$. In the asymptotic limit $r \to \infty$ each radial parameter converges to its appropriate constant value, as in Eqs.(4.40) and (4.41). For example, $lim_{r\to\infty}\alpha_s(r) = \alpha_s$, and $lim_{r\to\infty}C_c(r) = -1$. In addition, when r becomes much larger than the physical extent of the ionic core, $r_>$ coincides with the radial coordinate of the Rydberg electron, and the radial parameters

Table 6.1: $e - Mg^{++}$ model potential parameters.

	ℓ	α_1^ℓ	α_2^ℓ	α_3^ℓ	r_c^ℓ
Mg^{++}	0	4.51367	11.81954	2.97141	1.447764
$\alpha_d = 0.49$	1	4.71475	10.71581	2.59888	1.71333
	≥ 2	2.99158	7.69976	4.38828	1.730930
Ref.[71]					

assume their constant asymptotic values. Analytic expressions for these radial parameters are given in Appendix A.

6.2 Radial Dependence of Core Parameters

The accurate calculation of electronic properties for a many-electron ion is generally very difficult. Fortunately, Mg^+ is an alkaline-earth ion, for which simple and effective model potentials [70, 56, 71] can be utilized. The model potential $V(r)$ describing the interaction between a valence electron and a closed-shell ionic core is chosen to have the analytic form

$$V(r) = -\frac{1}{r}\left[2 + (Z-2)exp(-\alpha_1^\ell r) + \alpha_2^\ell r\ exp(-\alpha_3^\ell r)\right] - \frac{\alpha_d}{2r^4}\left[1 - exp\left(-\left(\frac{r}{r_c^\ell}\right)^6\right)\right], \quad (6.3)$$

where Z is the nuclear charge and α_d is the scalar dipole polarizability of the doubly charged positive ion [70, 56, 71]. The parameters α_i^ℓ and r_c^ℓ are empirically fitted to obtain agreement between the energy eigenvalues of the one-electron model Hamiltonian and the experimental energy levels of the alkaline-earth ion. The values of the parameters for the interaction of Mg^{++} and a valence electron are given in Table 6.1. As noted in Ref.[70, 56] the ℓ dependence of these parameters makes the model potential formally a nonlocal potential, although in a trivial way.

The core wavefunctions are computed by diagonalizing the model Hamiltonian in a finite basis set of radial Sturmian functions (see Appendix C)

$$S_{n\ell}^{(\zeta)}(r) = \left[\frac{(n-\ell-1)!}{(n+\ell)!}\right]^{1/2} e^{-\zeta r/2}(\zeta r)^{\ell+1} L_{n-\ell-1}^{(2\ell+1)}(\zeta r), \tag{6.4}$$

where the $L_{n-\ell-1}^{(2\ell+1)}(\zeta r)$ are associated Laguerre polynomials defined by

$$L_n^{(k)}(x) = \sum_{\nu=0}^{n} \binom{n+k}{n-\nu} \frac{(-x)^\nu}{\nu!}, \tag{6.5}$$

and ζ is a parameter chosen to enhance convergence. Expanding each desired radial wavefunction as a linear combination of these Sturmians, the radial Schrödinger equation becomes a generalized eigenvalue problem

$$H_{model}\psi = EO\psi. \tag{6.6}$$

H_{model} and O are matrices of the model Hamiltonian and overlap operators, calculated in the Sturmian basis [110]. The asymptotic permanent multipole moments, polarizabilities, and hyperpolarizabilities for the $Mg^+(3p)$ ionic core computed with this wavefunction information are presented in Table 6.2.

A comparison with our earlier work on the Ne^+ ion reveals that the Mg^+3p ionic parameters are from one to two orders of magnitude larger. In particular, the vector hyperpolarizability β_v is $1.885 a.u.$ for the $Mg^+(3p)$, as opposed to $0.059(2) a.u.$ for the ground state Ne^+ ionic core [30, 31]. Unfortunately, as we will demonstrate later, the large autoionization widths of the $3pnf$ Rydberg states obscure the importance of this subtle interaction.

The channel independent radial parameters are also computed with this wavefunction information and are presented in Figure 6.1. In order to emphasize their radial variation, we have multiplied each parameter by a suitable constant. They converge rapidly to their asymptotic values, i.e. reasonably well by $r \sim 10$ a.u.. Nevertheless, the variations for $r < 10$ a.u. are sufficiently strong to shift the $3pnf$ Rydberg levels of magnesium with $n \sim 10$ by a few wavenumbers. For states with higher ℓ values, the centrifugal barrier excludes the

Table 6.2: Mg^+ $3p$ ionic core parameters.

Parameters	Calculated (a.u.)
C_c	-1
Q	2.752
α_s	31.514
α_t	1.531
η_s	677.205
η_t	-583.541
β_s	84.418
β_v	1.885
β_t	-34.144

Rydberg electron from the small radial region where these short range variations play an important role. These calculations were accurately converged, to $0.001 a.u.$ for each parameter, with 100 radial Sturmian functions per channel.

6.3 Rydberg Levels and Rates: Mg $3pnf$

The interactions among Rydberg channels generally increase in importance as the ionic core becomes larger and more polarizable. In this case the $3pnf$ Rydberg states of magnesium would appear ideal for the study of anisotropic interactions like $\frac{\vec{L}_c \cdot \vec{\ell}}{r^6}$. However, the small energy shifts generated by this vector interaction are difficult to observe in these states, because of the large autoionization widths. This could, in principle, be partially resolved by studying high ℓ states of a given n so that the decay rate is dominated by radiative decay, which is independent of n and ℓ [72]. In fact, this is preferable provided the natural linewidth for the $3p$ state is smaller than the splittings and energy shifts caused by the anisotropic terms.

Fortunately, the $3pnf$ states of Mg are still interesting, since strong coupling arises from the quadrupole interaction. This coupling gives rise to irregular behavior in the positions and the decay widths of the autoionization resonances, which can be clearly seen in the recent experiments of Lyons *et.al.* [32]. In fact, the coupling is strong enough that traditional perturbative treatments of Rydberg states are unable to account for the observed effects.

We compute the magnesium $3pnf$ Rydberg energy levels and the corresponding wavefunction information by diagonalizing the generalized eigenvalue equation, Eq.(3.8), with the computed radial functions shown in Figure 6.1. Once again we use a radial Sturmian basis for convenience. The autoionization rates (widths) are computed using the T-matrix formulation developed in Section 3.12. Since we exclusively consider Rydberg states with total angular momentum $J = 4$, only g final states contribute to these calculated rates.

The fine structure splitting between the Mg^+ $3p_{\frac{1}{2}}$ and $3p_{\frac{3}{2}}$ ionization thresholds was taken to be $91.57 cm^{-1}$ [73, 74]. The Rydberg energy levels, rates, and quantum defects

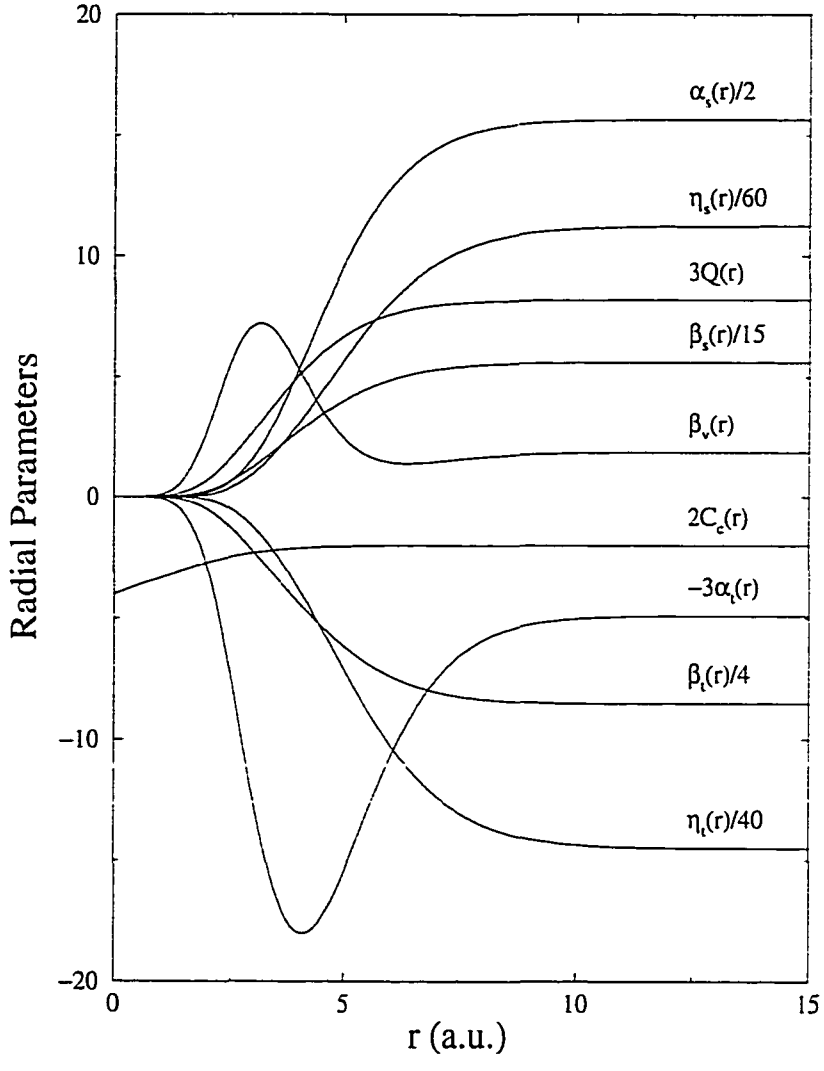


Figure 6.1: Channel-independent radial parameters for the Mg^+ $3p$ ion.

for the Mg $3p_{\frac{1}{2}}nf_{\frac{7}{2}}(J=4)$, $3p_{\frac{3}{2}}nf_{\frac{7}{2}}(J=4)$, and $3p_{\frac{3}{2}}nf_{\frac{9}{2}}(J=4)$ resonances are presented in Tables 6.3 through 6.5. The agreement between our computed and the measured levels is good to about $1cm^{-1}$. Since the measured rates have uncertainties of 20% and range from 5 to $10cm^{-1}$, depending on the symmetry, this agreement between our theory and experiment is reasonably good. The theoretical errors are primarily due to the neglect of exchange, which is of borderline importance for f states, and the use of a model potential to compute electronic properties of the Mg^+ $3p$ ion.

The irregular behavior of the $3p_{\frac{1}{2}}14f_{\frac{7}{2}}$ and $3p_{\frac{1}{2}}18f_{\frac{7}{2}}$ states, and of the $3p_{\frac{3}{2}}14f_{\frac{7}{2}}$ and $3p_{\frac{3}{2}}17f_{\frac{7}{2}}$ states is clearly seen in the n-dependence of their autoionization rates. To better visualize the distribution of these rates, we plot them as functions of the principal quantum number n in Figures 6.2 through 6.4, respectively. The principal quantum number n and the effective quantum number n^* are related by the quantum defect (QD) μ ($n^* = n - \mu$).

Apart from the occasional perturbation, the general decrease of these rates with increasing n is consistent with the $\frac{1}{n^3}$ trend when $n \gg \ell$ [72]. The irregular behavior is caused by strong channel coupling, which is primarily due to the quadrupole interaction, and by the near degeneracies of the Rydberg levels in different channels, such as the $3p_{\frac{1}{2}}14f_{\frac{7}{2}}$ and the $3p_{\frac{3}{2}}13f_K$ states. For comparison, we also present in Figures 6.2 through 6.4, the results of an R-matrix, multichannel quantum defect theory (MQDT) calculation, which agrees well with the effective Hamiltonian approach. The R-matrix results were based on reaction matrices calculated (inside a reaction volume of $20a.u.$) in an earlier study [75].

This comparison between the effective Hamiltonian approach and the more traditional and well tested R-matrix method is revealing. In general the positions and widths computed using $\bar{H}^{PP}$ agree better with experiment. This may not be surprising since R-matrix methods restrict the coordinate space to a reaction volume that may not necessarily be large enough to describe the spatial distribution of high $n\ell$ Rydberg states. The Hamiltonian approach simply requires an increase in the number of radial Sturmian functions per

Table 6.3: Mg $3p_{\frac{1}{2}}nf_{\frac{7}{2}}J=4$ energies (cm^{-1}), widths (cm^{-1}), and quantum defects.

n	E expt. [32]	E theory	Γ expt. [32]	Γ theory	QD expt. [32]	QD theory
9	95967.83	95969.94	6.48	8.326	0.053	0.0457
10	96231.89	96231.30	5.60	6.534	0.042	0.0449
11	96424.39	96424.35	4.90	5.193	0.044	0.0441
12	96571.94	96571.01	3.69	4.223	0.036	0.0431
13	96686.31	96685.11	3.43	3.597	0.030	0.0411
14	96778.00	96776.80	4.95	3.627	0.009	0.0235
15	96848.72	96847.71	1.62	1.805	0.033	0.0480
16	96907.94	96907.48	1.55	1.709	0.037	0.0451
17	96957.11	96956.93	1.38	1.559	0.039	0.0430
18	96999.49	96999.07	3.60	1.798	0.010	0.0212
19	97033.17	97033.09	0.73	0.972	0.044	0.0466

Table 6.4: Mg $3p_{\frac{3}{2}}nf_{\frac{7}{2}}J=4$ energies (cm^{-1}), widths (cm^{-1}), and quantum defects.

n	E expt. [32]	E theory	Γ expt. [32]	Γ theory	QD expt. [32]	QD theory
9	96050.34	96052.63	3.63	3.898	0.082	0.0746
10	96314.82	96315.89	2.54	2.504	0.081	0.0761
11	96509.75	96510.33	1.19	1.636	0.081	0.0774
12	96658.25	96657.90	0.54	0.969	0.076	0.0794
13	96771.41	96771.34		0.00150	0.093	0.0938
14	96865.43	96864.23	1.48	1.114	0.060	0.0748
15	96938.06	96937.32	0.95	0.639	0.066	0.0779
16	96996.76	96996.41	0.37	0.000619	0.087	0.0938
17	97046.97	97046.97	1.19	0.505	0.076	0.0769
18	97088.3	97087.75	0.59	0.0159	0.082	0.0974
19		97123.63	0.66	0.295		0.0785

Table 6.5: Mg $3p_{\frac{3}{2}}nf_{\frac{5}{2}}J=4$ energies (cm^{-1}), widths (cm^{-1}), and quantum defects.

n	E expt. [32]	E theory	Γ expt. [32]	Γ theory	QD expt. [32]	QD theory
9	96065.66	96067.23	11.53	12.212	0.032	0.0269
10	96326.73	96326.72	9.25	9.053	0.027	0.0275
11	96518.26	96518.61	6.28	6.882	0.030	0.0279
12	96665.55	96664.49	5.26	5.349	0.020	0.0282
13	96778.31	96777.98	5.05	3.932	0.025	0.0282
14	96869.47	96867.96	3.07	3.405	0.010	0.0286
15	96941.58	96940.54	2.13	2.781	0.012	0.0288
16	97000.40	96999.93	2.64	2.264	0.020	0.0288
17	97049.01	97049.13	1.73	1.921	0.031	0.0290
18	97090.85	97090.35	2.15	1.615	0.015	0.0291
19		97125.22	2.36	1.382		0.0292

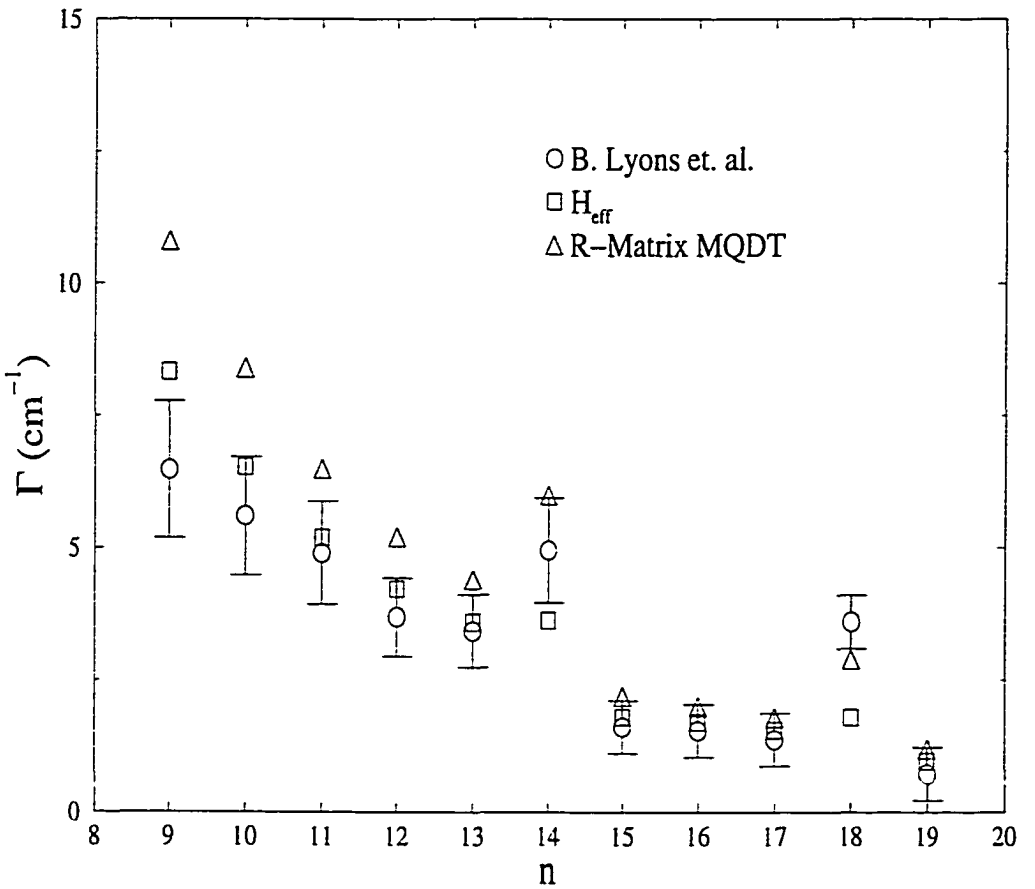


Figure 6.2. Plot of the Mg $3p_{\frac{1}{2}}nf_{\frac{7}{2}}(J=4)$ autoionization rates as a function of the principal quantum number n. The circled data points with error bars are experimental results [32].

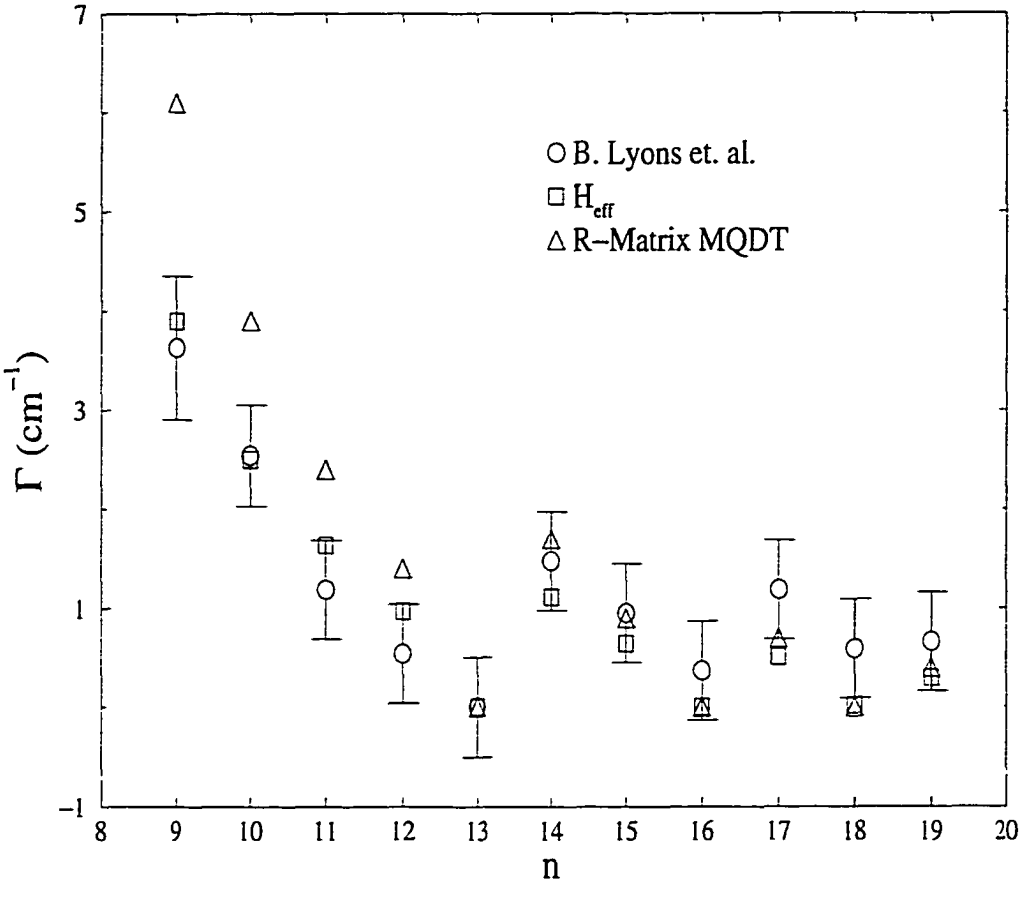


Figure 6.3. Plot of the Mg $3p_{\frac{3}{2}}nf_{\frac{7}{2}}(J=4)$ autoionization rates as a function of the principal quantum number n. The circled data points with error bars are experimental results [32].

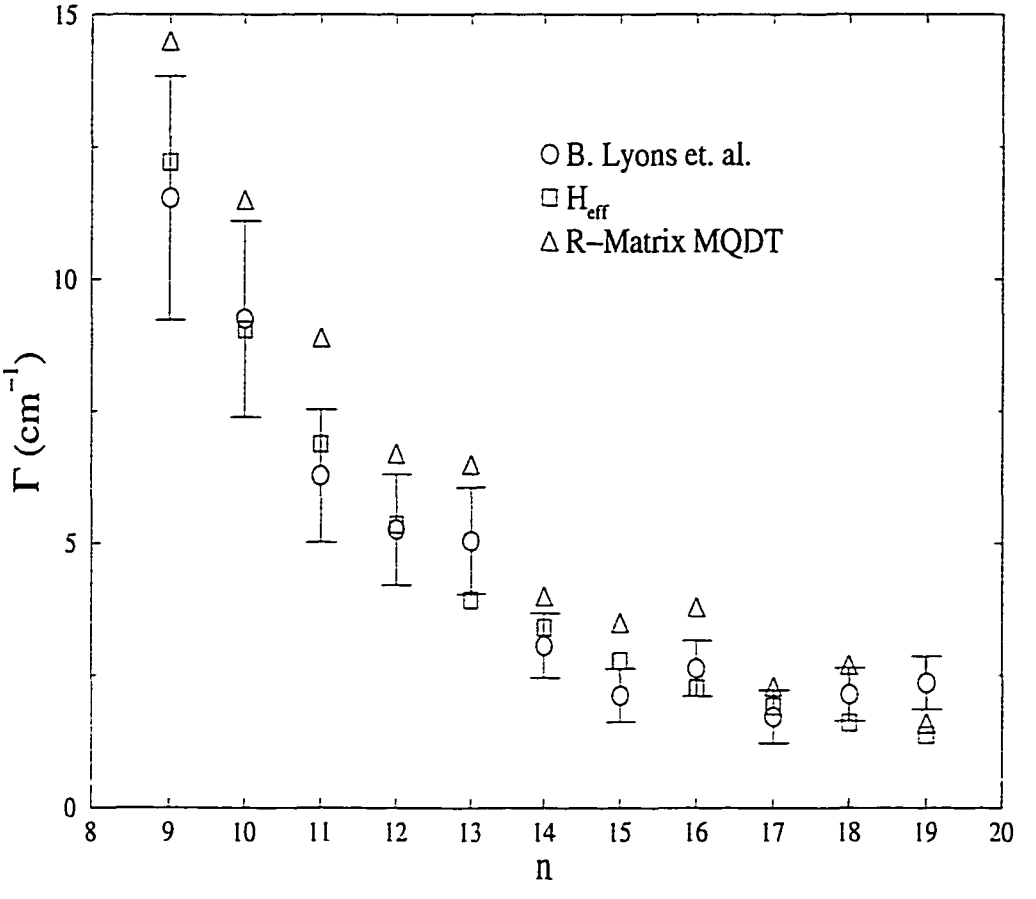


Figure 6.4. Plot of the Mg $3p_{\frac{3}{2}}nf_{\frac{5}{2}}(J=4)$ autoionization rates as a function of the principal quantum number n. The circled data points with error bars are experimental results [32].

channel to treat higher $n\ell$ states.

6.4 Adiabatic Torquing of Orbital Planes

In this section we more fully explore the idea that the ionic core of a Rydberg system experiences a torque from the distant, slow moving Rydberg electron. Here we turn to an adiabatic analysis to study how the orbital planes of the ionic core and the Rydberg electron orient themselves as the Rydberg electron adiabatically changes its radial position. Explicitly, the adiabatic potentials $U_\mu(r)$ and eigenstates $\Phi_\mu(r;\Omega)$ are defined as the r-dependent solutions of the generalized eigenvalue problem:

$$\bar{H}^{PP}_{r=const}\Phi_\mu(r;\Omega) = U_\mu(r)\Lambda^{PP}_{r=cont}\Phi_\mu(r;\Omega). \tag{6.7}$$

The adiabatic channel functions $\{\Phi_\mu(r;\Omega)\}$ are superpositions of the ionic core states and the orbital functions of the Rydberg electron; they contain information relating to the instantaneous interactions between the ionic core and the distant electron.

The $K^\pi = \frac{7}{2}^-$ and $K^\pi = \frac{9}{2}^-$ adiabatic potentials for Rydberg Mg, shown in Figures 6.5 and 6.6, exhibit very broad and smooth avoided crossings, and derivative couplings with similar qualitative behavior. In order to fit the potential curves and the couplings in the same figure we have divided the derivative couplings by a factor of twenty. The large magnitude of these couplings once again suggests that a diabatic treatment, where the coupling is still relatively small, is preferable. As a result of this coupling, Rydberg states attached to these potential curves should strongly interfere with one another and produce irregular behavior in the positions and decay widths of the autoionization resonances, as described in the previous section.

The orbital angular momentum of the Rydberg electron and that of the ionic core can be viewed as defining orbital planes as in classical mechanics. Here we investigate how the orientation of these orbital planes, or the projection of the Rydberg orbital momentum

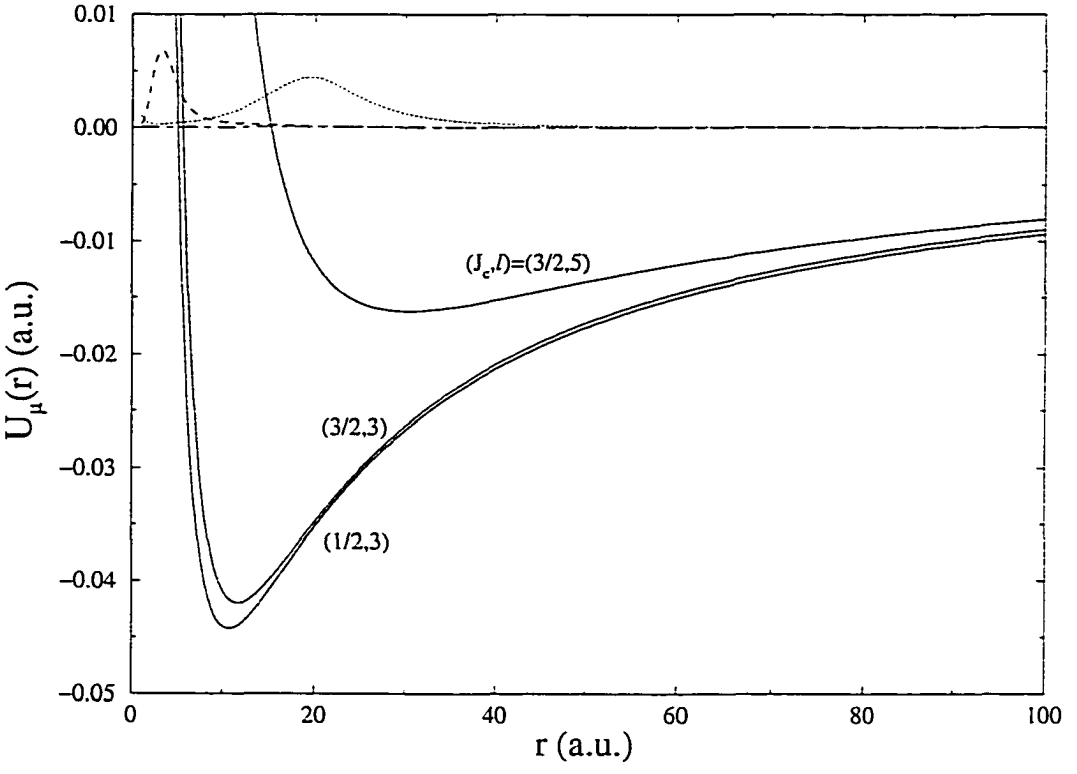


Figure 6.5. Radial adiabatic potential curves and derivative couplings for the $K^\pi = \frac{7}{2}^-$ Rydberg states of Mg. The derivative couplings $P_{\mu\mu'}(r)/20$ are given by the broken lines; $P_{bottom,middle}(r)$, dotted; $P_{middle,top}(r)$, dashed; and $P_{bottom,top}(r)$, dot-dashed.

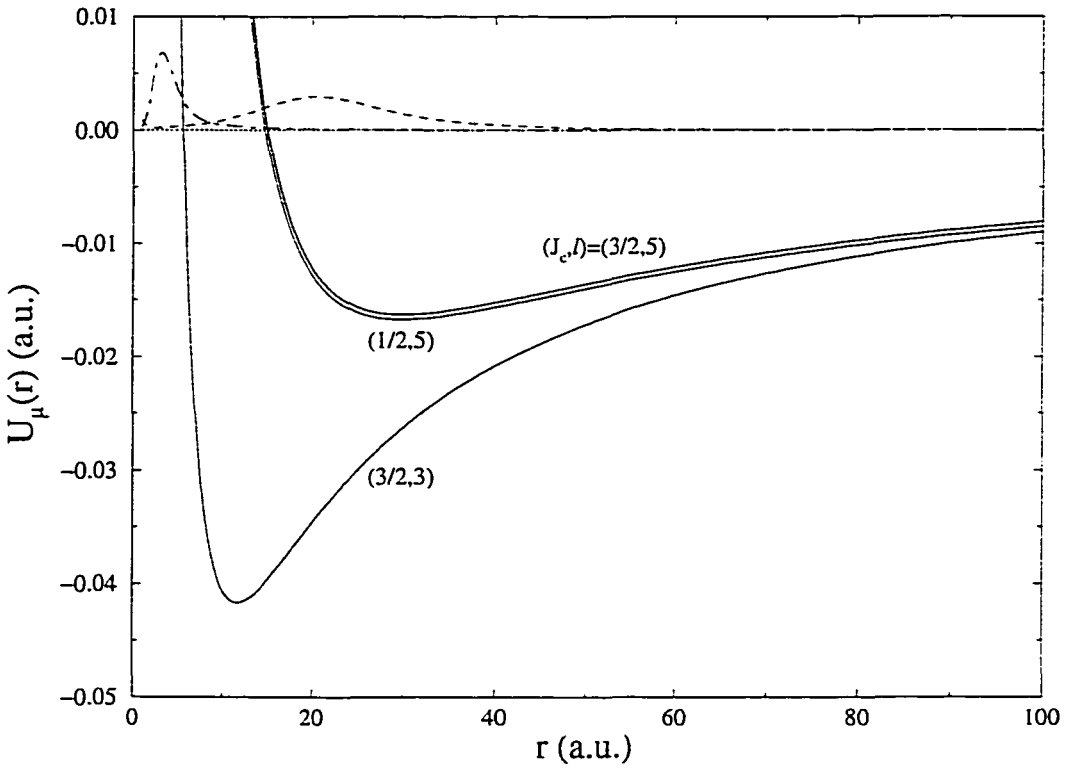


Figure 6.6. Radial adiabatic potential curves and derivative coupling for the $K^\pi = \frac{9}{2}^-$ Rydberg states of Mg. The derivative couplings $P_{\mu\mu'}(r)/20$ are given by the broken lines; $P_{bottom,middle}(r)$, dotted; $P_{middle,top}(r)$, dashed; and $P_{bottom,top}(r)$, dot-dashed.

onto the core orbital momentum, varies in a radial adiabatic analysis of Rydberg Mg. The matrix element of interest is:

$$\langle \vec{L_c} \cdot \vec{\ell} \rangle_\mu = \langle \Phi_\mu \mid \vec{L_c} \cdot \vec{\ell} \mid \Phi_\mu \rangle, \tag{6.8}$$

which varies with r for each of the $K^\pi = \frac{7}{2}^-$ and $K^\pi = \frac{9}{2}^-$ channels of Rydberg Mg.

In Figure 6.7 we plot $\langle \vec{L_c} \cdot \vec{\ell} \rangle_\mu$ versus r for the three adiabatic channels $(J_c, \ell) = (\frac{1}{2}, 3)$, $(\frac{3}{2}, 3)$, and $(\frac{3}{2}, 5)$ of $K^\pi = \frac{7}{2}^-$ Mg. In the isolated channel $(J_c, \ell) = (\frac{3}{2}, 5)$ the relative orientation of $\vec{\ell}$ and $\vec{L_c}$ exhibits very little variation with r. In contrast, the orientation of these two vectors shows a dramatic variation in the $(J_c, \ell) = (\frac{1}{2}, 3)$ and $(J_c, \ell) = (\frac{3}{2}, 3)$ channels. In these cases the variation reflects the presence of an avoided crossing that can be attributed to the strong coupling between these two channels. One channel of particular interest is $(J_c, \ell) = (\frac{1}{2}, 3)$ where the relative orientation flips, or changes sign. We attribute this type of variation in orientation to the anisotropic interactions and to the spin-orbit coupling within the system, since without the additional spin angular momentum $\vec{L_c} \cdot \vec{\ell}$ would be conserved.

In Figure 6.8 we show another example where anisotropic interactions in combination with core spin-orbit coupling produce interesting effects. Here we plot $\langle \vec{L_c} \cdot \vec{\ell} \rangle_\mu$ versus r, but for the three adiabatic channels $(J_c, \ell) = (\frac{1}{2}, 5)$, $(\frac{3}{2}, 3)$, and $(\frac{3}{2}, 5)$ of $K^\pi = \frac{9}{2}^-$ Mg. Once again, the isolated channel $(J_c, \ell) = (\frac{3}{2}, 3)$ shows very little effect from the radial variation. However, the two remaining channels exhibit a crossing, but no sign reversals. For large radial distances the relative orientation of the core and Rydberg angular momenta remains relatively fixed. As the Rydberg electron approaches the ionic core their electrostatic interaction increases, causing the relative orientation of the two orbital planes to change, provided the core is anisotropic.

In these two examples, the orbital plane of the ionic core undergoes the largest reorientation. This can be understood by realizing that total angular momentum is conserved,

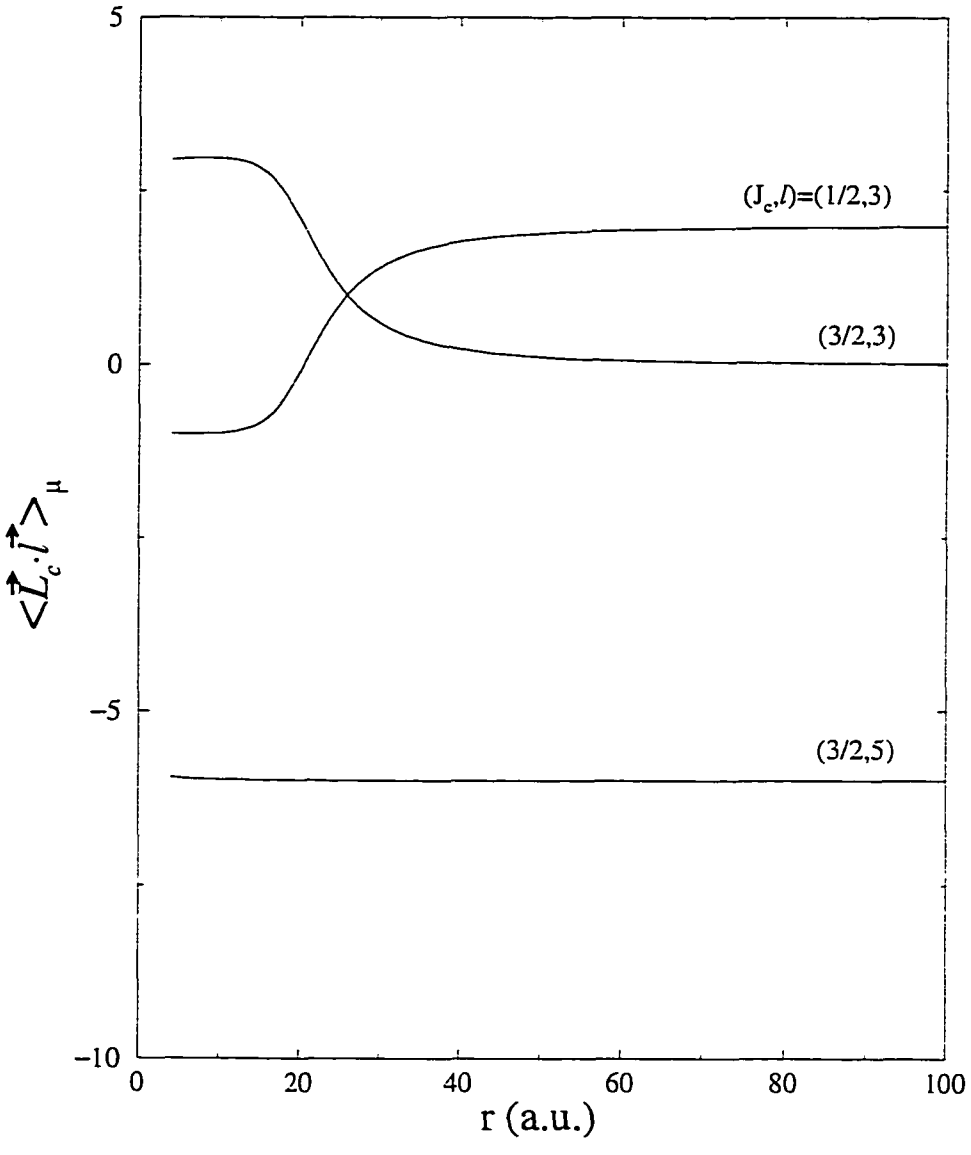


Figure 6.7. Adiabatic variation in the orientation of the ionic core and the Rydberg orbital planes for $K^\pi = \frac{7}{2}^-$ Rydberg Mg

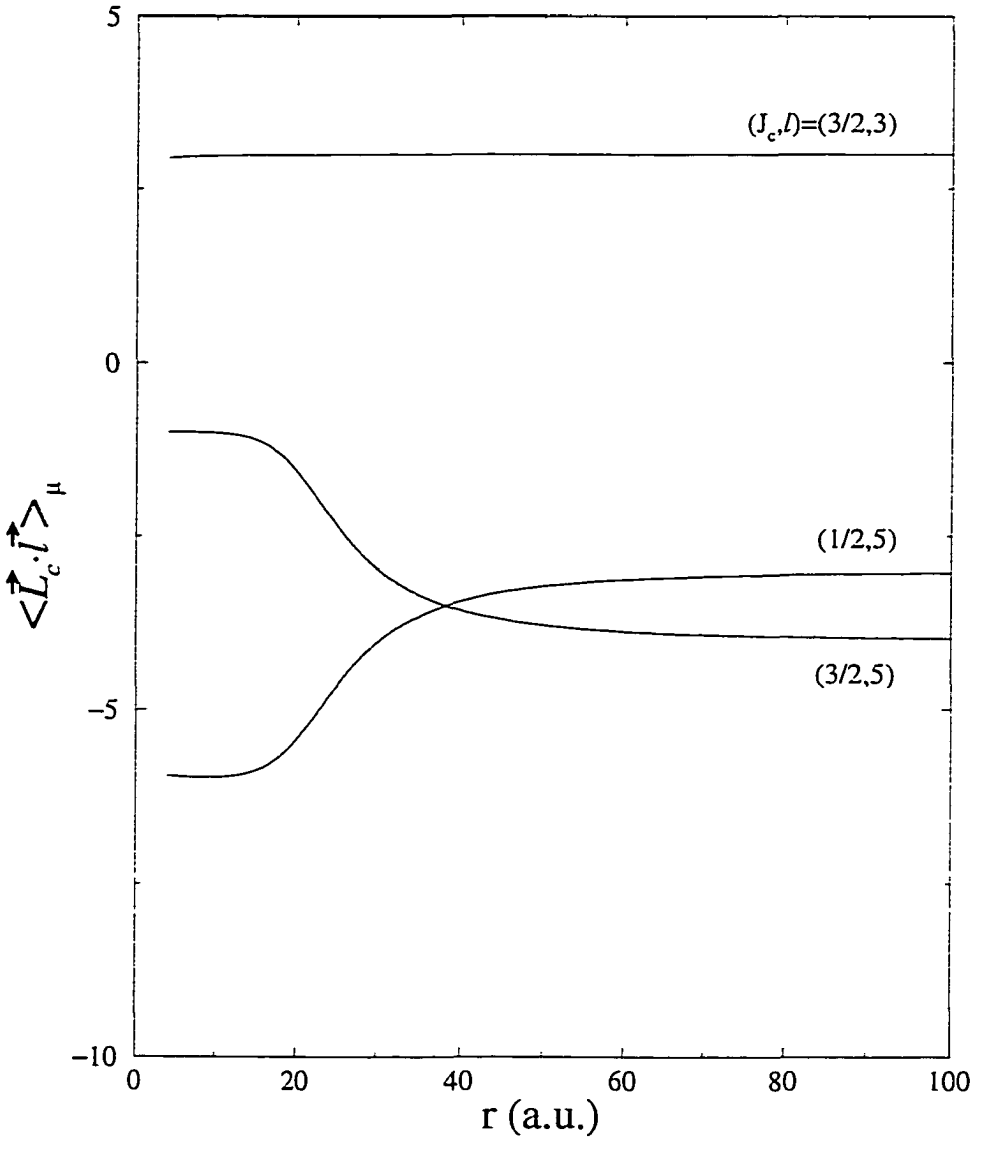


Figure 6.8. Adiabatic variation in the orientation of the ionic core and the Rydberg orbital planes for $K^\pi = \frac{9}{2}^-$ Rydberg Mg

and, therefore, only the much smaller angular momentum associated with the ionic core can exhibit a sign change. In this sense, we can view the orbit of the ionic core as being "flipped" by the approaching Rydberg electron.

CHAPTER 7

RYDBERG STATES OF THE HYDROGEN DIATOMIC MOLECULE

In this chapter we present our study on Rydberg states of H_2 and D_2. The recent high resolution measurements of Rydberg intervals in ground rovibrational state H_2 and D_2 [9] contain interesting physics for several reasons. The simplicity of the H_2^+ and D_2^+ ionic cores enables theory to compute all of the electric multipole and the induced polarization interactions with enough precision to be competitive with the best experimental measurements. Moreover, these measurements are tantalizingly close to resolving, for the first time, the slight energy shifts of Rydberg states due to relativistic retardation. The tiny interaction associated with retardation in this context is sometimes called a "Casimir force" [9, 19]. Excitement generated by this experimental progress has renewed theoretical interest in Rydberg state physics and in precise calculations of H_2^+ and D_2^+ ionic properties. This chapter develops the most extensive, complete adiabatic treatment of H_2^+ and D_2^+ ion properties presented to date. The resulting comparison with experiment gives a stringent test of our diabatic formulation of the Rydberg states physics.

7.1 Molecular Hamiltonian in Jacobi Coordinates

We begin by developing a general theory of the interactions in the generalized hydrogen Rydberg molecule. Much other work on molecular Rydberg systems has of course been done elsewhere [76, 77, 78, 79, 80, 81, 82, 11, 83, 84]. The primary reason for presenting this development here is to detail the differences among various approximations currently being used, and to clarify the advances our treatment contributes to the field of Rydberg state physics.

The formulation of a many-body problem in terms of relative coordinates gives

rise to an expression for the kinetic energy of the system that involves scalar products of momentum operators associated with different pairs of constituent particles; such terms have been referred to as "mass polarization" [86, 87, 88]. In almost all atoms and small molecules the mass polarization can be treated as a perturbation. However, it is desirable to absorb the perturbative effects of mass polarization into an effective Hamiltonian that describes how the Rydberg electron interacts with the ionic core. One way of doing this is through the use of Jacobi coordinates [89, 90, 11] where the position of each successive particle is defined relative to the center of mass of the previous subsystem.

7.1.1 Transformation Rydberg states of the generalized hydrogen diatomic molecule consist of two nuclei with charges $Z_a e$ and $Z_b e$, a core electron, and a Rydberg electron. The laboratory positions of each particle, respectively, are defined as $\vec{r}_a$, $\vec{r}_b$, $\vec{r}_i$, and $\vec{r}_{Ryd}$ relative to a common origin. The transformation from $\{\vec{r}_a, \vec{r}_b, \vec{r}_i, \vec{r}_{Ryd}\}$ to the Jacobi coordinates $\{\vec{R}_{cm}, \vec{R}, \vec{r}_e, \vec{r}\}$ is advantageous, in that no mass polarization terms arise in the kinetic energy operator. The center of mass of the entire molecule is defined by

$$\vec{R}_{cm} = \frac{M_a \vec{r}_a + M_b \vec{r}_b + m_e \vec{r}_i + m_e \vec{r}_{Ryd}}{M_a + M_b + 2m_e}. \tag{7.1}$$

The relative position of the two nuclei is

$$\vec{R} = \vec{r}_a - \vec{r}_b. \tag{7.2}$$

The position of the core electron relative to the center of mass of the nuclei is

$$\vec{r}_e = \vec{r}_i - \frac{M_a \vec{r}_a + M_b \vec{r}_b}{M_a + M_b}, \tag{7.3}$$

and

$$\vec{r} = \vec{r}_{Ryd} - \frac{M_a\vec{r}_a + M_b\vec{r}_b + m_e\vec{r}_i}{M_a + M_b + m_e} \tag{7.4}$$

is the position of the Rydberg electron relative to the center of mass of the molecular ion.

7.1.2 Kinetic and Potential Operators In Jacobi coordinates the kinetic (T) and potential (V) operators of a generalized hydrogen diatomic molecule are given by

$$\begin{aligned} T \quad &= \frac{1}{2(M_a + M_b + 2m_e)}\vec{P}^2_{cm} + \frac{M_a + M_b}{2M_aM_b}\vec{P}^2_R \\ &+ \frac{M_a + M_b + m_e}{2m_e(M_a + M_b)}\vec{P}^2_e + \frac{M_a + M_b + 2m_e}{2m_e(M_a + M_b + m_e)}\vec{P}^2_r, \end{aligned} \tag{7.5}$$

and

$$\begin{aligned} V \quad &= \frac{Z_aZ_be^2}{R} - \frac{Z_ae^2}{|\,\vec{r}_e - \epsilon_a\vec{R}\,|} - \frac{Z_be^2}{|\,\vec{r}_e + \epsilon_b\vec{R}\,|} \\ &- \frac{Z_ae^2}{|\,\vec{r} - \epsilon_a\vec{R} + \epsilon\vec{r}_e\,|} - \frac{Z_be^2}{|\,\vec{r} + \epsilon_b\vec{R} + \epsilon\vec{r}_e\,|} + \frac{e^2}{|\,\vec{r} - \vec{r}_e + \epsilon\vec{r}_e\,|}. \end{aligned} \tag{7.6}$$

The mass factors in the potential energy, which would normally be associated with mass polarization when non-Jacobian type coordinates are use, are

$$\epsilon_a \equiv \frac{M_b}{M_a + M_b}, \qquad \epsilon_b \equiv \frac{M_a}{M_b + M_a}, \qquad \epsilon \equiv \frac{m_e}{M_a + M_b + m_e}. \tag{7.7}$$

Here ϵ_a and ϵ_b depend only on the masses of the two nuclei, while ϵ also depends on the mass of the electron.

7.1.3 Molecular Ion and Rydberg Hamiltonians Since we are interested in high-ℓ Rydberg states of H_2 and D_2, the full molecular Hamiltonian can be decomposed into largely separate molecular ion and Rydberg electron Hamiltonians:

$$H = T + V = H_{core} + H_{Rydberg}. \tag{7.8}$$

All terms relating to the Rydberg electron, including the electron-ion interaction, are included in the latter. The Hamiltonian for the molecular ion is

$$\begin{aligned} H_{core} \quad &\equiv \frac{M_a + M_b}{2M_aM_b}\vec{P}_R^2 + \frac{Z_aZ_be^2}{R} \\ &+ \frac{M_a + M_b + m_e}{2m_e(M_a + M_b)}\vec{P}_e^2 - \frac{Z_ae^2}{\mid \vec{r}_e - \epsilon_a\vec{R} \mid} - \frac{Z_be^2}{\mid \vec{r}_e + \epsilon_b\vec{R} \mid}, \end{aligned} \tag{7.9}$$

and the Hamiltonian for the Rydberg electron is

$$\begin{aligned} H_{Rydberg} \quad &\equiv \frac{M_a + M_b + 2m_e}{2m_e(M_a + M_b + m_e)}\vec{P}_r^2 + \frac{e^2}{\mid \vec{r} - \vec{r}_e + \epsilon\vec{r}_e \mid} \\ &- \frac{Z_ae^2}{\mid \vec{r} - \epsilon_a\vec{R} + \epsilon\vec{r}_e \mid} - \frac{Z_be^2}{\mid \vec{r} + \epsilon_b\vec{R} + \epsilon\vec{r}_e \mid}. \end{aligned} \tag{7.10}$$

Here it is important to note that H_{core} is the Hamiltonian for a free molecular ion, independent of any Rydberg coordinate. However, $H_{Rydberg}$ is coupled to the molecular ion through its dependence on the internuclear separation $\vec{R}$ and the position $\vec{r}_e$ of the molecular electron. These two Hamiltonians do not commute, of course.

7.1.4 Spherical Expansion of the Rydberg Hamiltonian As in atomic Rydberg systems, the large value of the Rydberg radial coordinate facilitates a perturbative expansion of all electrostatic interactions between the Rydberg electron and the molecular ion. After an expansion of the Rydberg-core interaction into spherical multipoles, the Rydberg Hamiltonian simplifies to

$$H_{Rydberg} \equiv \frac{1}{2}\frac{M_a + M_b + 2m_e}{m_e(M_a + M_b + m_e)}\vec{P}_r^2 + e^2\frac{1 - Z_a - Z_b}{r}$$

$$+e^2\sum_{k>0}\left[N_k r_e^k C^{(k)}(\hat{r}_e) - M_k R^k C^{(k)}(\hat{R})\right]\cdot\frac{C^{(k)}(\hat{r})}{r^{k+1}} \quad (7.11)$$

where the mass factors are

$$N_k \equiv (1-\epsilon)^k - (-1)^k(Z_a + Z_b)\epsilon^k, \quad (7.12)$$

and

$$M_k \equiv Z_a\epsilon_a^k + (-1)^k Z_b\epsilon_b^k. \quad (7.13)$$

For Rydberg states of H_2 this Hamiltonian (in atomic units based on the bare, not reduced, electron mass) becomes

$$H_{Rydberg} \equiv -\frac{1}{2}\frac{2M_p + 2}{2M_p + 1}\vec{\nabla}_r^2 - \frac{1}{r}$$

$$+\sum_{k>0}\left[N_k r_e^k C^{(k)}(\hat{r}_e) - M_k R^k C^{(k)}(\hat{R})\right]\cdot\frac{C^{(k)}(\hat{r})}{r^{k+1}}. \quad (7.14)$$

Here M_p is the proton mass and $m_e = 1$. For this case, N_k and M_k reduce to the simple expressions

$$N_k \equiv (1-\epsilon)^k - 2(-1)^k\epsilon^k, \quad (7.15)$$

and

$$M_k \equiv \left(\frac{1}{2}\right)^k \left(1 + (-1)^k\right) . \tag{7.16}$$

To treat Rydberg states of D_2 we need only change the masses in the molecular ion and Rydberg Hamiltonians. The nuclear mass factor M_k is nonzero only when k is *even*.

7.2 Close-Coupling Representation

The theoretical description of a Rydberg diatomic molecule is essentially the same as that for a Rydberg atom, except for the additional vibrational (ν^+) and rotational (N^+) structure. The wavefunction for the entire molecular system can be expanded in a primitive representation $\{\phi_i(\omega)\}$ constructed from the energy eigenstates of the molecular ion and the angular functions of the Rydberg electron

$$\Psi(r,\omega) = \sum_i \phi_i(\omega)\psi_i(r) \tag{7.17}$$

where ω represents all coordinates in the molecular system except the Rydberg radial position. Since we are still considering nonpenetrating, high-ℓ Rydberg states, for which electron spin effects (such as exchange) are negligible, the appropriate coupling scheme is still $(N^+\ell)K$ coupling [35, 36], when molecular hyperfine and spin-rotation splittings can be neglected.

The rotational structure of the molecular ion plays a role analogous to the fine structure of the atomic ions considered ealier. The close-coupling equations once again provide a useful channel structure that facilitates the separation of physically dominant channels from those that play only a perturbative role in the description of a Rydberg molecule. Following the spirit of Chapter 3, we use the standard technique of channel elimination [46] and a Green's function expansion [49, 50] to derive a long-range effective diabatic Hamiltonian that describes how the distant Rydberg electron interacts with the molecular ion. A tensorial analysis similar to that described in Chapter 4 for Eq.(4.40) generates an effective Hamiltonian with operator structure

$$
\begin{aligned}
H_{\mu\mu'} \quad &= \left(E_\mu - E_{\mu_0} - \frac{1}{2\mu_m}\frac{d^2}{dr^2} + \frac{\ell_\mu(\ell_\mu+1)}{2\mu_m r^2} - \frac{1}{r} - \frac{\alpha_s}{2r^4} - \frac{\eta_s}{r^6}\right)\delta_{\mu\mu'} \\
&- \left(\frac{Q}{r^3} + \frac{\alpha_t}{2r^4} + \frac{\eta_t}{r^6}\right)\mathcal{P}^{(2)}_{\mu\mu'} - \left(\frac{\phi}{r^5} + \frac{\lambda_h}{r^6}\right)\mathcal{P}^{(4)}_{\mu\mu'} \\
&+ \left(\frac{\beta_s \mathcal{B}^{(0)}_{\mu\mu'} + \beta_t \mathcal{B}^{(2)}_{\mu\mu'}}{r^6}\right) + \frac{\beta_v}{r^6}\langle \vec{N^+}\cdot\vec{\ell}\rangle_{\mu\mu'} \\
&- 2\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{P}^{(2)}_{\mu\mu'}}{r^5}\right) - \frac{d}{dr}\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{P}^{(2)}_{\mu\mu'}}{r^4}\right)\frac{d}{dr}
\end{aligned}
\tag{7.18}
$$

where E_{μ_0} is the lowest ionization threshold in the physical channel subspace, and $\mu_m = \frac{2M_p+1}{2M_p+2}$ is the reduced mass of the Rydberg electron relative to the ion. Likewise, the $\Lambda_{\mu\mu'}$ matrix in Eq.(4.41) is given by

$$
\Lambda_{\mu\mu'} = \delta_{\mu\mu'} + 2\left(\frac{\beta_s\delta_{\mu\mu'} + \beta_t \mathcal{P}^{(2)}_{\mu\mu'}}{r^4}\right). \tag{7.19}
$$

Here we adopt the convention in which α_s and α_t are the standard scalar and tensor induced dipole polarizabilities, Q and ϕ are the permanent electric quadrupole and hexadecapole moments, and η_s, η_t, and λ_h are higher-order scalar, second-rank tensor, and fourth-rank tensor induced hyperpolarizabilities of the ionic core. Explicit expressions for all terms, including the angular factors $\langle \vec{N^+}\cdot\vec{\ell}\rangle_{\mu\mu'}$, $\mathcal{P}^{(2)}_{\mu\mu'}$, and $\mathcal{P}^{(4)}_{\mu\mu'}$ are given in Appendix B.

As in the atomic case, the β terms represent nonadiabatic effects arising from the radial and angular motion of the Rydberg electron. Here the vector interaction involving β_v couples the total angular momentum N^+ of the molecular ion with the orbital angular momentum ℓ of the Rydberg electron. Interestingly, since the rovibrational states necessarily involve both the massive nuclei and the molecular electron, effects arising from the vector interaction will be much smaller in molecular systems than in atomic systems. In essence,

it is more difficult for a Rydberg electron to "drag" the polarization vector of a massive spinning molecular ion, as opposed to the polarization vector of an atomic ion that involves much lighter electrons. We will return to this point in a couple of sections.

7.3 Levels of Approximation

In order to compute Rydberg state properties of diatomic hydrogen in our formulation, we must determine the permanent electric multipole moments and the induced polarizabilities and hyperpolarizabilities of the molecular ion. The calculation of ionic core properties is a difficult task, but fortunately there are a number of familiar molecular approximations that can be utilized. These approximations make use of the fact that massive nuclei move far more slowly than the molecular electrons.

The wavefunction Ψ_{ion} for the molecular ion can be expanded in products of rovibrational and electronic wavefunctions

$$\Psi_{ion}(\vec{R},\vec{r_e}) = \sum_j F_j(\vec{R})\Phi_j(\vec{R},\vec{r_e}), \tag{7.20}$$

where the electronic wavefunction $\Phi_j(\vec{R},\vec{r_e})$ satisfies the equation

$$\left(-\frac{1}{2}\frac{2M_p+1}{2M_p}\vec{\nabla}_e^2 - \frac{1}{\mid \vec{r_e}-\frac{1}{2}\vec{R}\mid} - \frac{1}{\mid \vec{r_e}+\frac{1}{2}\vec{R}\mid}\right)\Phi_j(\vec{R},\vec{r_e}) = E_j^e(R)\Phi_j(\vec{R},\vec{r_e}). \tag{7.21}$$

An equation for the rovibrational wavefunction $F_i(\vec{R})$ of the nuclei is obtained after projecting out the electronic states in the full molecular ion Schrödinger equation

$$\left(-\frac{1}{M_p}\vec{\nabla}_R^2 + \frac{1}{R} + E_i^e(R)\right)F_i(\vec{R}) - 2\frac{1}{M_p}\sum_j\left[\langle\phi_i \mid \vec{\nabla}_R \mid \phi_j\rangle\cdot\vec{\nabla}_R\right]F_j(\vec{R})$$

$$-\frac{1}{M_p}\sum_j\left[\langle\phi_i \mid \vec{\nabla}_R^2 \mid \phi_j\rangle\right]F_j(\vec{R}) = E_{core}F_i(\vec{R}). \tag{7.22}$$

Various methods can be used to solve these equation, ranging from the Born-Oppenheimer approximation to exact, nonadiabatic methods which treat all coordinates including the internuclear separation R on an equal basis [78, 79, 80, 81, 82, 91, 92, 93, 94, 95, 96, 97, 85]. For clarity, let us review the two most familiar approximate methods that provide solutions for the rovibrational equation: the Born-Oppenheimer approximation and the adiabatic-nuclei approximation.

7.3.1 Born-Oppenheimer Approximation In the Born-Oppenheimer approximation the nuclei are assumed initially to be infinitely massive. The relative separation between the nuclei then becomes a constant parameter. The nuclear momentum operator $-i\vec{\nabla}_R$ is set to zero in this initial step, in Eq.(7.22) leaving

$$\left(-\frac{1}{M_p}\vec{\nabla}_R^2 + \frac{1}{R} + E_i^e(R)\right) F_i^{BO}(\vec{R}) = E_{core}^{BO} F_i^{BO}(\vec{R}) \tag{7.23}$$

as the nuclear rovibrational equation, which contains no coupling among different rovibrational states. The extremely useful Born-Oppenheimer approximation has provided much of the framework used today to understand the low vibrational states of diatomic molecules. However, this approximation must be greatly improved upon to account for modern, high-resolution Rydberg state spectroscopy, which can probe the fine details of the interactions among electronic, vibrational, and rotational degrees of freedom.

7.3.2 Adiabatic Nuclei Approximation The adiabatic-nuclei approximation assumes that the distance between the nuclei varies slowly, but is not entirely negligible compared to the motion of the molecular electrons. In this case Eq.(7.22) reduces to

$$\left(-\frac{1}{M_p}\vec{\nabla}_R^2 + \frac{1}{R} + E_i^e(R) - \frac{1}{M_p}\langle\phi_i \mid \vec{\nabla}_R^2 \mid \phi_i\rangle\right) F_i^{AD}(\vec{R}) = E_{core}^{AD} F_i^{AD}(\vec{R}) \tag{7.24}$$

where the diagonal matrix elements involving $\vec{\nabla}_R$ vanish, since the matrix is skew-symmetric, but the matrix elements involving $\vec{\nabla}_R^2$ survive as corrections due to the relative motion of

the nuclei. While coupling among different rovibrational states is not included at this level of approximation, the energy levels of low rovibrational states are generally accurate to about five significant figures [91, 85]. As we will demonstrate later in this chapter, the adiabatic-nuclei approximation can account for most the physics observed in the recent high-resolution experimental studies of H_2 and D_2 [9].

In our studies of Rydberg states of H_2 and D_2 the rovibrational state of the molecular ion in the physical channel subspace is usually either the ground state or the first rotationally excited state. By the Frank-Condon principle, only excited states that significantly overlap with these low lying states, which have near zero nuclear kinetic energy, will contribute to the matrix elements in the summations for the polarizabilities and hyperpolarizabilities [77, 36]. For this reason, the adiabatic-nuclei approximation can be used to accurately compute low rovibrational induced polarizabilities and hyperpolarizabilities of H_2^+ and D_2^+.

7.4 Core States of H_2^+ and D_2^+

All of our calculations for the H_2^+ and D_2^+ ions are performed using the adiabatic-nuclei approximation. These adiabatic calculations of the induced polarizabilities and hyperpolarizabilities, which involve complete summations over rovibrational states of these molecular ions, are the first of their kind. In almost all previous calculations of these quantities, a "completeness" approximation has been used [98, 99, 100, 101, 19].

7.4.1 Calculation of Electronic States The solution of the electronic equation Eq.(7.21) is most easily performed in prolate spheroidal coordinates (see Appendix D) where the partial differential equation breaks into three separate equations whose solutions are linked only through separation constants. In prolate spheroidal coordinates the electronic Hamiltonian (using atomic units with the bare electron mass) is [78, 79, 80, 81, 82, 77]

$$H_{elec} = -\frac{2}{R^2(\xi^2-\eta^2)}\left\{\frac{\partial}{\partial\xi}(\xi^2-1)\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\eta}(1-\eta^2)\frac{\partial}{\partial\eta}+\frac{(\xi^2-\eta^2)}{(\xi^2-1)(1-\eta^2)}\frac{\partial^2}{\partial\phi^2}\right\}$$

$$-\frac{4\xi}{R}\frac{1}{\xi^2-\eta^2}. \tag{7.25}$$

After electronic wavefunction is expanded as a product

$$\Phi(R,\vec{r_e}) = L(R,\xi)M(R,\eta)N(\phi), \tag{7.26}$$

the electronic equation separates into

$$\left[\frac{\partial}{\partial\xi}(\xi^2-1)\frac{\partial}{\partial\xi}-A+\frac{ER^2}{2}\xi^2+R\xi(Z_a+Z_b)-\frac{\lambda^2}{\xi^2-1}\right]L(R,\xi)=0 \tag{7.27}$$

$$\left[\frac{\partial}{\partial\eta}(1-\eta^2)\frac{\partial}{\partial\eta}+A-\frac{ER^2}{2}\eta^2-R\eta(Z_a-Z_b)-\frac{\lambda^2}{1-\eta^2}\right]M(R,\eta)=0 \tag{7.28}$$

$$-i\frac{\partial}{\partial\phi}N(\phi)=\lambda N(\phi) \tag{7.29}$$

where λ and A are separation constants and E is the electronic energy of the system. Note that A is a function of λ and E.

The azimuthal equation can be solved analytically, with the result

$$N(\phi)=\frac{1}{\sqrt{2\pi}}e^{i\lambda\phi} \tag{7.30}$$

where $\lambda = 0, \pm1, \pm2, \cdots$. Some of the symmetry properties of the hydrogen diatomic molecular ion are immediately apparent from these equations. Since λ only appears in the

form λ^2 there is a double degeneracy for all nonzero λ states, which is consistent with the azimuthal-symmetry of the electronic Hamiltonian about the internuclear axis. Furthermore, for homonuclear diatomic ions, where $Z_a = Z_b$, the $Z_a - Z_b$ term in the angular $M(R,\eta)$ equation vanishes and the angular equation becomes invariant under reflection ($\eta \rightarrow -\eta$). This invariance translates into angular eigenfunctions that possess either gerade (even) or ungerade (odd) symmetry for each value of the separation constant A.

Solution of the pair of radial $L(R,\xi)$ and angular $N(R,\eta)$ equations requires a little more effort since they are coupled through the separation constant A and the electronic energy E. While there are many ways of solving these equations [78, 79, 80, 81, 82], an efficient and accurate method is to solve each equation using a finite element method (see Appendix E), and iterate the solutions until A and E converge to some desired precision, which in our calculations is ten digits.

7.4.2 Calculation of Rovibrational States The rovibrational states of the molecular ion are obtained by solving the adiabatic-nuclei equation Eq.(7.24), with the electronic energies $E_i^e(R)$, which depend on the internuclear separation R, and the adiabatic corrections arising from the $\vec{\nabla}_R^2$ nuclear operator. A finite element method (see Appendix E) is used to solve these equations. The energy eigenvalues obtained are converged to ten digits, while the core electric multipole moments and the induced polarizabilities and hyperpolarizabilities are converged to only about five digits. (Variational eigenfunctions are typically accurate to only half as many digits as the quantity that is variational, which is the energy in this case.)

7.4.3 Core Parameters and the Vector Hyperpolarizability Until this study, no complete adiabatic calculation of any polarizability or hyperpolarizability had been performed for H_2^+ or D_2^+. The best previous calculations were based on what is loosely called a "clamped nuclei" approximation [98, 99, 100, 101, 19] that neglects the small rovibrational splittings in the energy denominators of the induced polarizabilities and hyperpolarizabilities.

A closure relationship for the rovibrational states can then be used to eliminate the summations over intermediate rovibrational quantum numbers, thereby leaving simple (purely electronic) polarizabilities and hyperpolarizabilities. As with the Born-Oppenheimer approximation, the "clamped nuclei" approximation is unacceptable in the present context, owing to the high resolution attained in Rydberg state spectroscopy.

A few "exact", or nonadiabatic calculations of the permanent electric multipole moments have been performed by Bishop and Moss[102, 93]. However, the first nonadiabatic calculation of the ground state dipole polarizability $\alpha_s = 3.1682(4)$ of H_2^+ has been completed only very recently by Shertzer [103]. This value is still unpublished and should be regarded as preliminary as of this writing. Nevertheless, the new value agrees well with the indirect measurement $\alpha_s = 3.1682(7)$ of Ref.[9]. Thus far, Shertzer's method has only been employed to compute the dipole polarizabilities of the ground rovibrational states of H_2^+ and D_2^+. An extension of Shertzer's method to other ground state or excited rovibrational state polarizabilities and hyperpolarizabilities is likely to be technically difficult, and results are not expected to emerge for some time. For now, our calculation is the most complete treatment of the multipole and induced polarization interactions in H_2 and D_2. It is also the first calculation of the vector hyperpolarizability for a molecular system.

Table 7.1 shows our adiabatic calculation of the ground rovibrational state polarizabilities and hyperpolarizabilities for H_2^+ and D_2^+. These results were obtained by solving the electronic and rovibrational equations using the adiabatic-nuclei approximation. Summations were included over numerous rovibrational states of the molecular ions, until it appeared the series had converged. A total of 180 electronic states of H_2^+ and D_2^+ were computed out to $R = 8a.u.$ to ten digit accuracy. For each electronic potential and each rotational state, 200 rovibrational wavefunctions were computed using the adiabatic-nuclei approximation, also to ten digit accuracy.

Any residual errors in our scalar dipole polarizabilities α_s for H_2^+ and D_2^+, should

be caused by effects of nonadiabatic nuclear motion. In fact, the 0.046 percent difference between the exact and the adiabatic dipole polarizabilities is consistent with the estimates of Bishop [102] and Moss [93] who predicted a 0.043 percent difference between the exact and adiabatic quadrupole moments of $(\nu^+ = 0, N^+ = 1)$ H_2^+. While there are slight differences between the adiabatic calculations and the measured polarizabilities, the adiabatic calculations represent a significant improvement on the "clamped nuclei" approximation which predicts a scalar dipole polarizability $\alpha_s = 3.1730$, with a -0.15 percent difference relative to the exact calculation of Shertzer [103].

In the future we hope to study Rydberg states attached to the anisotropic $(\nu^+ = 0, N^+ = 1)$ H_2^+ ion. With this in mind, we present the first complete adiabatic calculation of the permanent electric multipole moments, polarizabilities and hyperpolarizabilities for this ion in Table 7.2. To our initial surprise, the vector hyperpolarizability β_v is an order of magnitude smaller for the $(\nu^+ = 0, N^+ = 1)$ H_2^+ ion than for the Ne^+ ion studied in Chapter 5. Moreover, β_v is three orders of magnitude smaller than β_s and β_t. Since the H_2^+ molecular ion is generally more polarizable than the Ne^+ ion, we initially expected the vector hyperpolarizability to be much larger, and to play a more significant role in the physics of H_2 Rydberg states. A study of this unexpected result has helped to elucidate the qualitative meaning of the vector interaction.

The explanation lies in the rotational structure of the molecular ion. The adiabatic-nuclei definition of the vector hyperpolarizability given in Appendix B is

$$\beta_{v_\mu} = (-1)^{N^+} N_1^2 \sqrt{\frac{3}{2}} \left[\frac{(2N^+ + 1)}{N^+(N^+ + 1)}\right]^{\frac{1}{2}}$$

$$\sum_\nu (-1)^{N_\nu^+} \frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2} \begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix} \begin{pmatrix} N_\nu^+ & 1 & N^+ \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

Table 7.1. Theoretical and experimental [9] parameters (atomic units) for ground state H_2^+ and D_2^+. The theoretical values were computed using a complete summation over adiabatic rovibrational states method, without making the usual "completeness approximation".

H_2^+	α_s	β_s	η_s
Adiabatic Theory	3.1667	1.6938	12.5601
Experiment [9]	3.1681(7)		
Nonadiabatic [103]	3.1682(4)		
Clamped Nuclei Approx. [85]	3.1730		
D_2^+	α_s	β_s	η_s
Adiabatic Theory	3.0708	1.5930	11.8807
Experiment [9]	3.0712(7)		
Nonadiabatic [103]	3.0714(4)		
Clamped Nuclei Approx. [9]	3.0739		

$$\left\{ \begin{array}{ccc} 1 & 1 & 1 \\ N^+ & N^+ & N_\nu^+ \end{array} \right\} \langle \Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu \rangle \langle \Psi_\nu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\mu \rangle^*. \tag{7.31}$$

Using the "clamped nuclei" approximation we can binomially expand the energy factor in powers of the vibrational and rotational energy splittings, which are far smaller than the electronic energy differences. Specifically,

$$\frac{1}{(\Delta E_{\mu_0 \nu})^2} = \frac{1}{(\Delta E^{el}_{\mu_0 \nu})^2} - \frac{2\Delta E^{vib}_{\mu_0 \nu}}{(\Delta E^{el}_{\mu_0 \nu})^3} - \frac{2\Delta E^{rot}_{\mu_0 \nu}}{(\Delta E^{el}_{\mu_0 \nu})^3} + \cdots. \tag{7.32}$$

where $\Delta E_{\mu_0 \nu} \equiv E_{\mu_0} - E_\nu$. The electronic ($el$) and vibrational ($vib$) energy differences in this expression contribute nothing to β_v, at this level of approximation, since the summation over the rotational momenta N_ν^+ vanishes. The rotational energy difference on the other hand survives, and provides our desired approximation to leading order. If we replace the rotational energy difference with

$$\Delta E^{rot}_{\mu_0 \nu} = \frac{[N^+(N^+ + 1) - N_\nu^+(N_\nu^+ + 1)]}{2I}, \tag{7.33}$$

where I is the moment of inertia of the diatomic ion, then the vector hyperpolarizability can be expressed in the form

$$\beta_{v_\mu} = \frac{\beta^{\parallel}_{v_\mu} + \beta^{\perp}_{v_\nu}}{2I}. \tag{7.34}$$

Here $\beta^{\parallel}_{v_\mu}$ and $\beta^{\perp}_{v_\nu}$ are parallel and perpendicular electronic hyperpolarizabilities that involve cubed energy denominators. Explicitly

$$\beta^{\parallel}_{v_\mu} = \int d\vec{R} F^*_\mu(\vec{R})\, \beta^{\parallel}_v(R)\, F_\mu(\vec{R}), \tag{7.35}$$

and

$$\beta^{\perp}_{v_\mu} = \int d\vec{R} F^*_\mu(\vec{R})\ \beta^{\perp}_v(R)\ F_\mu(\vec{R}), \tag{7.36}$$

where $F_\mu(\vec{R})$ is an adiabatic rovibrational wavefunction. The radial $\beta^{\parallel}_v(R)$ and $\beta^{\perp}_v(R)$ factors depend only on electronic properties of the molecular ion:

$$\beta^{\parallel}_v(R) = -2N_1^2 \sum_\alpha \frac{D^*_{X,\alpha 0} D_{X,\alpha 0}}{(E_X - E_{\alpha 0})^3}, \tag{7.37}$$

and

$$\beta^{\perp}_v(R) = -2N_1^2 \sum_\alpha \frac{D^*_{X,\alpha 1} D_{X,\alpha 1}}{(E_X - E_{\alpha 1})^3}, \tag{7.38}$$

where N_1 is the electronic mass factor defined in Eq.(7.15), and where the electronic dipole matrix element is defined as

$$D_{X,\alpha\lambda}(R) \equiv \int d\vec{r}_e \Phi^*_X(R,\vec{r}_e) r_e C^{(1)}_{-\lambda}(\hat{r}_e) \Phi_{\alpha\lambda}(R,\vec{r}_e). \tag{7.39}$$

The subscipts X ($\alpha\lambda$) denote the electronic ground (excited) states of H_2^+, and both E_X and $E_{\alpha\lambda}$ depend on the internuclear separation R.

The dependence of the β_{v_μ} expression on the moment of inertia of the molecular ion immediately explains why β_v is approximately one-thousand times smaller than the β_s and β_t hyperpolarizabilities. Moreover, the fact that there is a moment of inertia in this expression for β_v confirms our physical picture of the vector interaction. The origin of the vector interaction is qualitatively different from that of the static polarizabilities. The vector interaction describes how the motion of the Rydberg electron couples with dynamics of the molecular ion. As the distant Rydberg electron revolves around the core, it tries to

"drag" the core polarization vector with it. The comparatively enormous moment of inertia of the molecule resists this drag strongly, thereby minimizing the effect of this "torque". Consequently the dynamic vector interaction is miniscule in molecular Rydberg states.

7.5 Rydberg States of H_2 and D_2

We turn now to the $n = 9$ and 10 Rydberg states of $(\nu^+ = 0, N^+ = 0)$ H_2 and D_2, which have recently been studied experimentally by Jacobson *et al* [9]. In anticipation of new experimental studies of high-ℓ Rydberg states attached to the anisotropic $(\nu^+ = 0, N^+ = 1)$ H_2^+ molecular ion [104], we also present a quantitative investigation of the vector interaction that helps to elucidate its qualitative nature. These results are compared with theoretical calculations of the energy shifts generated by relativistic retardation (Casimir) [19] carried out elsewhere.

7.5.1 Ryberg States of $(\nu^+ = 0, N^+ = 0)$ H_2 and D_2 The experimentally observed spin splittings in the high-ℓ, $n = 9$ and 10 Rydberg states of $(\nu^+ = 0, N^+ = 0)$ H_2 and D_2 were analyzed by the experimental group to extract spinless Rydberg transition frequencies [9, 11, 105]. For these $N^+ = 0$ Rydberg states the hyperfine interaction produces an overall scalar shift that makes no contribution to any of the observed Rydberg intervals. Exchange energies were determined to be less than 0.01 MHz for these high-ℓ Rydberg states. Moreover, refinements made by Jacobson *et al* have improved the signal to noise ratio by a factor of 15 compared with earlier experiments on H_2 [11].

The effective diabatic Hamiltonian in Eq.(7.18) includes all permanent multipole and induced polarization interactions out to radial order $\left(\frac{1}{r}\right)^6$ and out to second-order in the Green's function energy expansion. Since these expansions are finite in extent, care must be taken to ensure that high-order terms contribute a negligible amount to the Rydberg energy intervals under study. Table 7.3 displays the first order energy shifts (MHz) of the $n = 10$, $\ell = 5, 6, 7, 8$ and 9 hydrogenic states arising from various perturbative $\frac{1}{r}$ interactions. The proportionality coefficient is taken to be $1 a.u.$ so that these results can be easily rescaled by

Table 7.2. Theoretical and experimental [11] parameters (atomic units) for $(v^+ = 0, N^+ = 1)$ H_2^+. These parameters are defined in Appendix B and the long range potential is given in Eq.(7.18). The theoretical values were computed using a complete summation over adiabatic rovibrational states method, without making the usual "completeness approximation".

H_2^+	Q	ϕ	λ_h
Adiabatic Theory	1.64285	2.01966	0
Experiment [11]	1.64323(30)		
	α_s	β_s	η_s
Adiabatic Theory	3.17627	1.70230	12.61887
Experiment [11]	3.1787(34)		
Clamped Nuclei Approx. [85]	3.1826		
	α_t	β_t	η_t
Adiabatic Theory	2.67900	2.07061	16.96614
Experiment [11]	2.678(24)		
Clamped Nuclei Approx. [85]	2.6863		
		β_v	
Adiabatic Theory		0.0048231	

the actual coefficients for any Rydberg system. Since our calculation of ionic core properties extends only to inverse radial powers of $\frac{1}{r^6}$, we can accurately predict Rydberg level intervals for ℓ-states for which higher-order $\frac{1}{r^7}$ and $\frac{1}{r^8}$ interactions are negligible. Experimental fits performed by Jacobson *et al* indicate that the coefficient of the scalar $\frac{1}{r^7}$ hyperpolarization interaction is approximately $20 a.u.$. On this basis, Table 7.3 indicates that contributions from this $\frac{1}{r^7}$ interaction and higher-order terms will be negligible for states with $\ell \geq 6$.

Using the theoretical core parameters from Table 7.1, we compute the $n = 9$ and 10 Rydberg intervals (and $\ell > 5$) for the ground rovibrational state of H_2^+ and D_2^+. The results of these calculations are presented in Table 7.4. The agreement between our calculations and the measured intervals is very good, considering the fact that we apply the adiabatic-nuclei approximation to compute all the induced polarizabilities and hyperpolarizabilities. While we do not explicitly compare with the predictions of a "clamped nuclei" calculation in Table 7.4, it is important to know that the error in the polarizabilities and hyperpolarizabilities predicts Rydberg intervals that are larger than the measured intervals by approximately 2 MHz for the $I - K$ transitions. Our calculations reproduce the experimental intervals with an accuracy of 0.3 MHz or better. While this accuracy is still not quite high enough to extract energy shifts due to relativistic retardation from these experiments, it suggests that further improvements in the calculations of the polarizabilities and hyperpolarizabilities of ground state H_2^+ and D_2^+ may clear up the remaining discrepancies, and hopefully improve our quantitative understanding of the non-retardation physics to the desired level.

7.5.2 Analysis of the Vector Interaction and Relativistic Retardation in the $n = 10$ Rydberg States of $(\nu^+ = 0, N^+ = 1)$ H_2 Since the first prediction of the vector interaction one of the most interesting questions in Rydberg state physics has been: what qualitative effect on the spectrum does the vector interaction generate? In order to better understand the nature of this unusual interaction, we compare the relative energy shifts generated by this term with those generated by the more typical even-rank tensorial

Table 7.3. Comparison of the $n = 10$, $\ell = 5, 6, 7, 8$ and 9 energy shifts (MHz) for various powers of $\frac{1}{r}$. Here the proportionality coefficient is taken to be 1.

Radial Dependence	ℓ	Energy Shift (MHz)
$\frac{1}{r^3}$	5	39876.87212000
	6	24101.40622637
	7	15665.91404714
	8	10751.11748333
	9	7695.53672491
$\frac{1}{r^4}$	5	1840.47102092
	6	753.71670381
	7	345.92606584
	8	172.01787973
	9	90.53572618
$\frac{1}{r^5}$	5	100.05735312
	6	27.38796162
	7	8.74266696
	8	3.07174785
	9	1.13169658
$\frac{1}{r^6}$	5	6.14464134
	6	1.10525641
	7	0.24316214
	8	0.05985970
	9	0.01508929
$\frac{1}{r^7}$	5	0.42749565
	6	0.04917661
	7	0.00735501
	8	0.00126020
	9	0.00021556
$\frac{1}{r^8}$	5	0.03447231
	6	0.00243449
	7	0.00024249
	8	0.00002864
	9	0.00000332

Table 7.4. Comparison of computed and measured [9] $n = 10$, $(v^+ = 0, N^+ = 0)$ Rydberg intervals for H_2 and D_2 (all results in MHz).

H_2			
Interval	ΔE^{obs}	ΔE^{theory}	theory-obs
$9I - 9K$	864.563(5)	864.683	0.120
$9K - 9L$	370.696(16)	370.931	0.235
$10I - 10K$	630.795(15)	630.731	-0.064
$10K - 10L$	274.072(20)	274.219	0.147
$10L - 10M$	129.673(20)	129.794	0.121

D_2			
Interval	ΔE^{obs}	ΔE^{theory}	theory-obs
$9I - 9K$	873.27(5)	873.445	0.175
$9K - 9L$	359.58(7)	359.895	0.315
$10I - 10K$	638.01(4)	638.100	0.090
$10K - 10L$	273.07(7)	273.065	-0.005
$10L - 10M$	129.436(14)	129.491	0.055

interaction terms, of second rank (quadrupole and tensor-polarizability) and fourth rank (hexadecapole).

Figure 7.1 displays the pattern of relative energy shifts generated by the operators in our effective long range potential, namely $\langle \vec{N}^+ \cdot \vec{\ell} \rangle_{\mu\mu}$, $\mathcal{P}^{(2)}_{\mu\mu}$, and $\mathcal{P}^{(4)}_{\mu\mu}$, as functions of the total angular momentum N of the system. Here the rotational angular momentum of the ionic core is taken to be $N^+ = 2$ and the orbital momentum of the Rydberg electron is fixed at $\ell = 6$. Each different rank operator expression produces a characteristic pattern of energy level splittings that can be observed experimentally. Specifically, the vector term produces a simple pattern known as the Landé interval rule in spectroscopy [86], with the splitting $E_N - E_{N-1}$ proportional to N. The lowest N state has the most negative shift, assuming the coefficient is positive, and each higher N state rises, which is qualitatively different from the patterns produced by the quadrupole and hexadecapole terms. In the quadrupole case we see a pattern where starting with the $N = 4$ state, which is positive, the next level $N = 5$ moves down just below zero, the next level $N = 6$ moves further down, and then the levels proceed back up to the $N = 7$ state, and finally up to the $N = 8$ state, which is positive. In the hexadecapole case the pattern alternates with four sign changes.

A comparison of the energy shifts produced by the vector interaction and by relativistic retardation is also revealing. Figure 7.2 compares the shifts generated by the vector interaction with the predicted Casimir shifts for the $n = 10$ Rydberg states of $(\nu^+ = 0, N^+ = 1)$ H_2 [19]. Here the molecular core is an anisotropic $(N^+ = 1)$ core and we consider three cases: $\ell = N$, $\ell = N + 1$, and $\ell = N - 1$. When the Rydberg orbital momentum ℓ is the same as the total angular momentum N of the system, or greater by one unit of angular momentum, we see that the vector shifts are negative, while the relativistic shifts are positive. In these two cases, a polarization model that does not include the vector interaction would predict Rydberg energy splittings that are slightly too large compared with experimental observations. When $\ell = N^+ - 1$ both the vector and relativistic shifts are

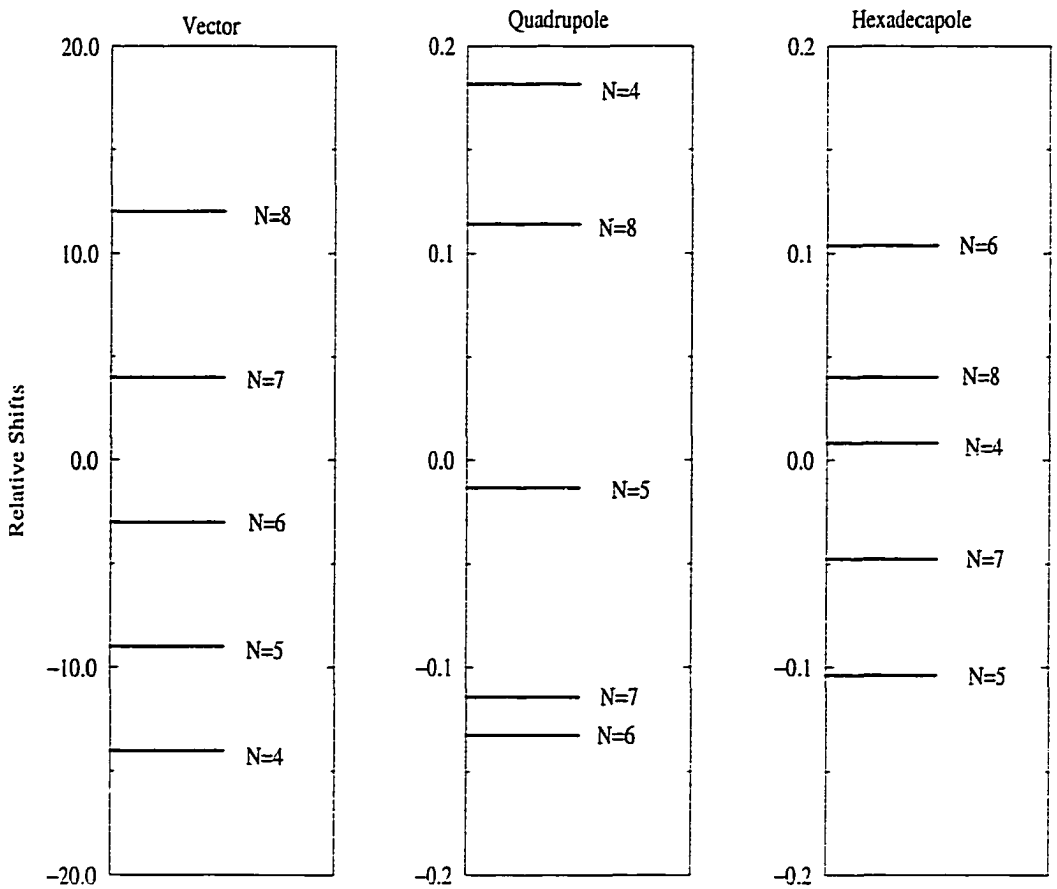


Figure 7.1. Comparison of relative shifts among vector, quadrupole, and hexadecapole operators versus the total angular momentum N. Here $N^+ = 2$ and $L = 6$.

positive, implying that neglect of the vector interaction would generate splittings that are too small.

Figure 7.3 also contrasts the tensorial structure of the vector and Casimir interactions. The N dependence of these two interactions is shown for the $n = 10$, fixed-ℓ Rydberg states of $(\nu^+ = 0, N^+ = 1)$ H_2. The distinct pattern of these energy shifts is similar to the shifts discussed earlier in Figure 7.1. These distinct tensorial patterns should stand out clearly in an experiment.

One additional feature to note is the more rapid decrease in the shifts with increasing ℓ produced by the vector interaction, as compared with those arising from retardation. The origin of this difference lies in their respective radial-dependences, which go as $\frac{1}{r^6}$ and $\frac{1}{r^5}$. Figure 7.4 displays this rapid decrease in the magnitudes of the vector energy shifts most clearly. Shifts due to the vector interaction dominate in the lower ℓ states, while the Casimir shifts dominate in higher ℓ states. Experimentalists aiming for an unambiguous confirmation of relativistic retardation in Rydberg atoms and molecules are advised, based on Figure 7.4, to focus on high $\ell > 6$ states. It would also greatly improve our understanding of the nonrelativistic effects to have measurements of far more energy splittings. The dependences on ℓ and N provide the most detailed information.

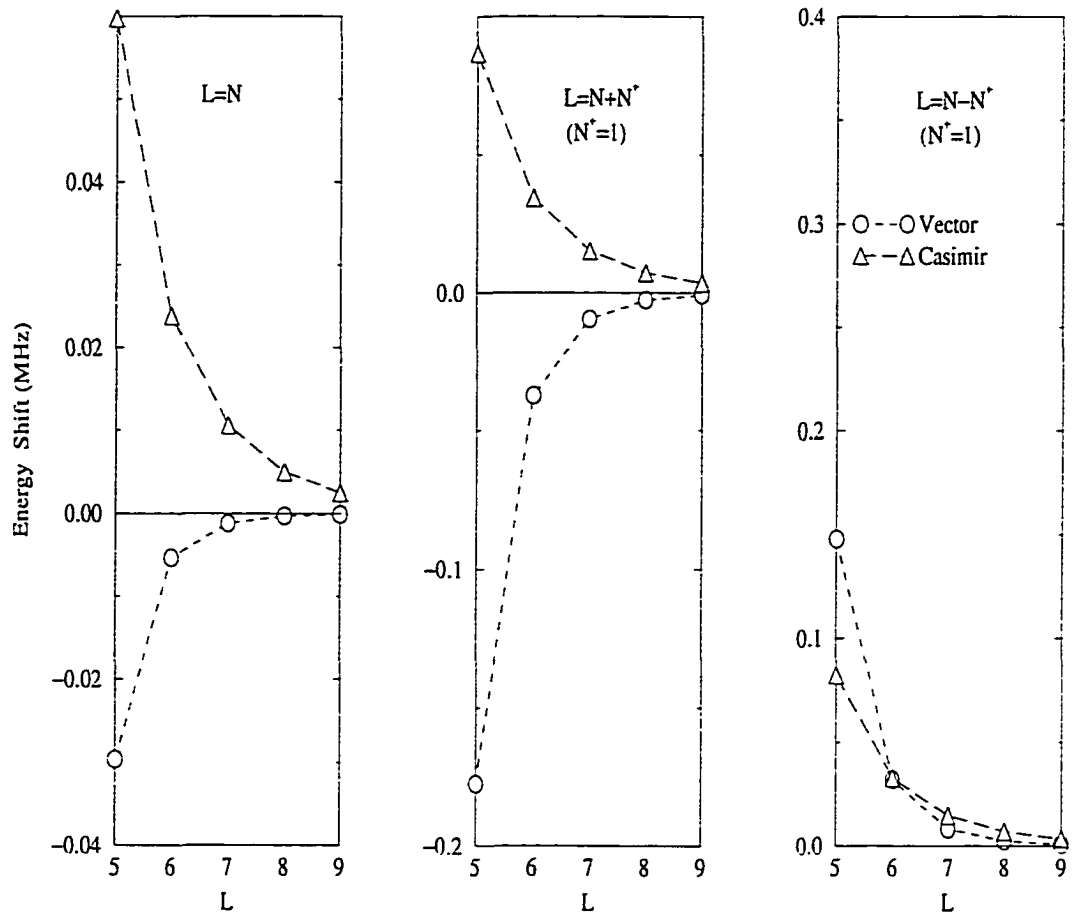


Figure 7.2. Comparison of the vector and Casimir energy shifts [19] in the $n = 10$, ($\nu^+ = 0, N^+ = 1$), $L = N$, $L = N - N^+$, and $L = N + N^+$ states of H_2.

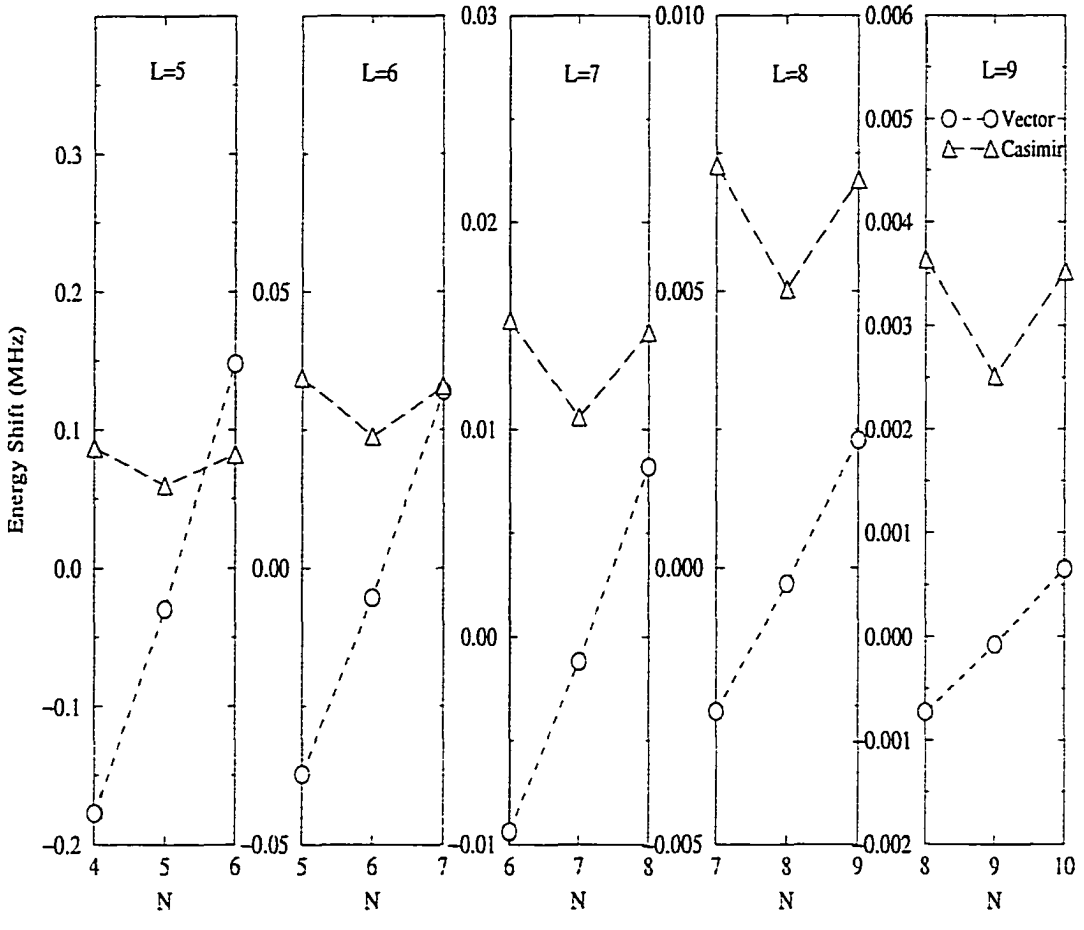


Figure 7.3. Comparison of the vector and Casimir energy shifts [19] in the $n = 10$, ($\nu^+ = 0, N^+ = 1$) states of H_2. Here the splittings for fixed Rydberg orbital momentum L are shown versus the total angular momentum N.

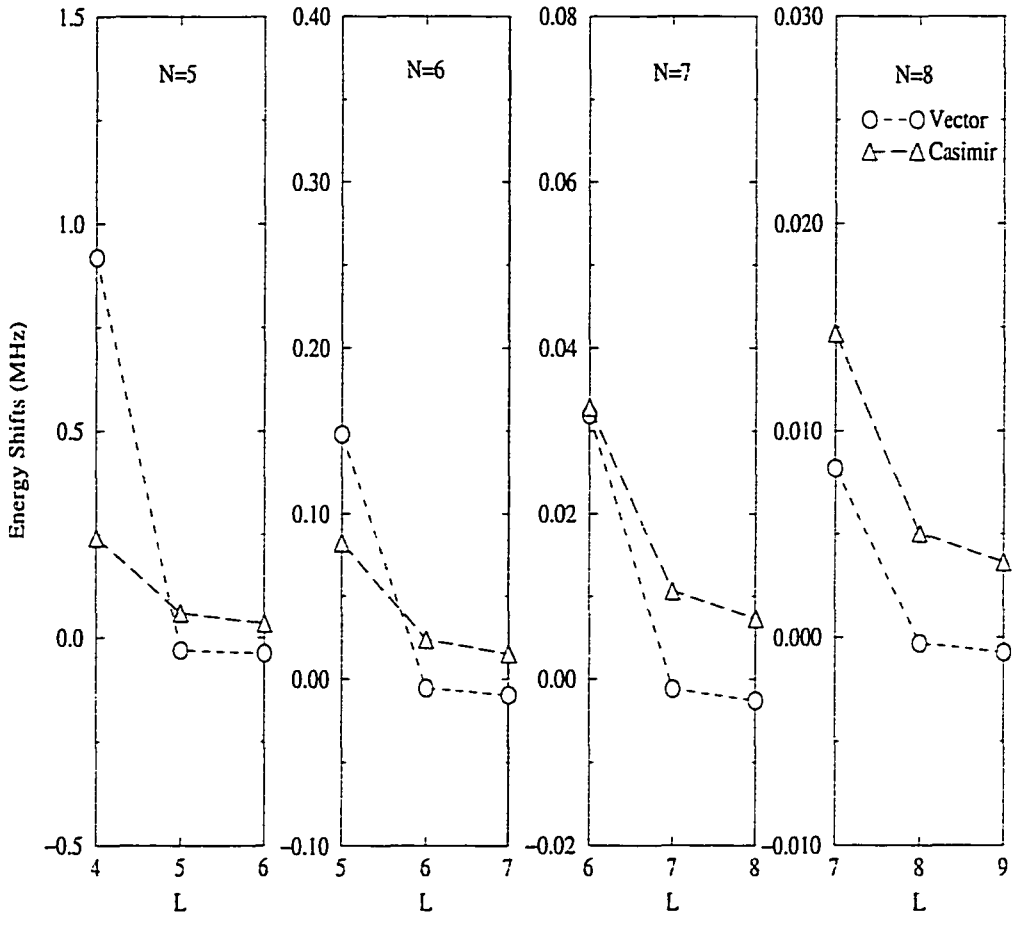


Figure 7.4. Comparison of the vector and Casimir energy shifts [19] in the $n = 10$, ($\nu^+ = 0, N^+ = 1$) states of H_2. Here the splittings for fixed total angular momentum N are shown versus the orbital angular momentum of the Rydberg electron L.

CHAPTER 8

SUMMARY AND REMARKS

In this dissertation, we have developed a new multichannel description of Rydberg state physics that combines the simplicity of long range effective potentials with the multichannel structure of modern scattering theory. At the same time, we have presented systematic adiabatic (and diabatic) derivations of an unusual vector interaction that has intrigued the Rydberg physics community in recent years. Our work demonstrates that the vector interaction describes a coupling among the orbital motion of the Rydberg electron and the internal ionic angular momentum, mediated by the dipole portion of the electron-ion Coulomb force. As the distant Rydberg electron revolves around the core, it tends to "drag" the core polarization vector with it. This drag is hindered by the internal moment of inertia of the ion, and by the moment of inertia of the distant electron (or other charged particle) about the center-of-mass. Stated another way, this interaction reflects the inability of the ionic core to instantaneously adjust to the angular motion of the Rydberg electron.

Our analysis of the $n = 10$ Rydberg states of Ne with orbital angular momenta of $\ell = 5, 6, 7$, and 8 confirmed, for the first time, the existence of an interaction term containing a pseudovector structure. We showed that this pseudovector interaction, in combination with an effective multichannel Hamiltonian, can reproduce observed Rydberg energy intervals in the $n = 10$ range to sub-MHz resolution. Moreover, a χ^2 comparison demonstrated that our description of Rydberg state physics improves on previously-developed perturbative polarization models. *Ab initio* calculations of theoretical values for $\alpha_s, \alpha_t, \beta_v, \cdots$ using multiconfiguration Hartree-Fock atomic wavefunctions confirm that the vector interaction exists and has experimental implications for Rydberg spectra with a distinct pattern of induced

energy splittings.

Our effective Hamiltonian approach was subsequently applied with success in our study of doubly excited, autoionizing $3pnf$ Rydberg states of Mg. The strong coupling among Rydberg states results in an irregular n-dependence of the autoionization rates, which was clearly seen in recent experiments and accurately described by our diabatic Hamiltonian approach. We also investigated variations of the relative orientation of the Rydberg and ionic core angular momenta in a radial adiabatic analysis of Rydberg Mg. The relative orientation of these angular momenta shows dramatic variations in strongly coupled channels, including a "flip", or sign change, in their mutual projection as the Rydberg electron distance from the nucleus changes. We attribute these variations in orientation to the combined effects of anisotropic interactions and the spin-orbit coupling within Mg^+.

Rydberg states of the simple diatomic molecules H_2 and D_2 served as a stringent test of our ability to compute ionic core properties, and to accurately describe Rydberg state physics. The theory and calculations presented are the most extensive, complete adiabatic treatment of H_2^+ and D_2^+ ion properties that has been presented to date. Our calculation of the permanent multipole moments and induced polarizabilities and hyperpolarizabilities for the first rotationally excited state of H_2^+ also provided a novel twist that has helped to elucidate the qualitative meaning of the vector interaction. In addition to being the first calculation of the vector hyperpolarizability for a molecule, the extremely small magnitude of the coefficient, one-thousand times smaller than other induced hyperpolarizabilities of H_2^+, proved to be enlightening. This analysis shows that the vector hyperpolarizability is inversely proportional to the moment of inertia of the diatomic ion. This connection confirms our physical picture of the vector interaction. As the distant Rydberg electron revolves around the core, it tries to "torque" the entire molecular ion. The relatively enormous rotational inertia strongly resists this drag, thereby making the dynamic vector interaction miniscule in Rydberg states of diatomic molecules. When the drag occurs more freely, as in an atomic

ion whose electrons have a very small moment of inertia, the energy associated with the drag shifts Rydberg levels more noticeably.

The $n = 9$ and 10 Rydberg energy intervals with $\ell > 5$ computed for the ground rovibrational states of H_2^+ and D_2^+ reproduce the experimental intervals with an accuracy of 0.3 MHz or better, a significant improvement on earlier calculations based on the "clamped nuclei" approximation. These adiabatic-nuclei calculations are not quite accurate enough yet to extract energy shifts due to relativistic retardation from these experiments. Future work will focus on computing higher order $\frac{1}{r^7}$ and $\frac{1}{r^8}$ hyperpolarization interactions and treating nuclear motion at a nonadiabatic level. An analysis of the energy shifts produced by the vector interaction and relativistic retardation demonstrate that these interactions generate clearly distinct patterns of energy level splittings, whose characteristic tensorial structure should stand out in an experiment. Experimentalists aiming for an unambiguous confirmation of relativistic retardation in Rydberg atoms and molecules should probably focus on high-ℓ Rydberg sates with $\ell > 6$. Moreover, we suggest that future experimental studies attempt to resolve far more energy level splittings, to permit the extraction of yet more information from the ℓ and N dependences of the Rydberg splittings.

BIBLIOGRAPHY

[1] W.E. Lamb and R.C. Retherford, Phys. Rev. **72**, 241 (1947).

[2] W.E. Lamb and R.C. Retherford, Phys. Rev. **79**, 549 (1950).

[3] H. Bethe, Phys. Rev. **72**, 339 (1947).

[4] H. Bethe, L.M. Brown, and J.R. Stehn, Phys. Rev. **77**, 370 (1950).

[5] R.P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); Phys. Rev. **74**, 939, 1430 (1948); Rev. Mod. Phys. **17**, 157 (1945).

[6] J. Schwinger, Phys. Rev. **73**, 416 (1948); Phys. Rev. **74**, 1439 (1948).

[7] S. Tomanaga, Prog. Theoret. Phys. **1**, 27 (1946); Prog. Theoret. Phys. **2**, 101, 198 (1947); Prog. Theoret. Phys. **3**, 1, 101 (1948); Phys. Rev. **74**, 224 (1948).

[8] Wood, C. S., S. C. Bennett, B. P. Masterson, J. L. Roberts, C. E. Tanner, C. E. Wieman, Science **275**, 1759 (1997).

[9] P.L. Jacobson, D.S. Fisher, C.W. Fehrenbach, W.G. Sturrus, and S.R. Lundeen, Phys. Rev. A **56**, R4361 (1997).

[10] R.F. Ward, Jr., W.G. Sturrus, and S.R. Lundeen, Phys. Rev. A **53**, 113 (1996).

[11] W.G. Sturrus, E.A. Hessels, P.W. Arcuni, and S.R. Lundeen, Phys. Rev. A **44**, 3032 (1991).

[12] W.G. Sturrus, E.A. Hessels, P.W. Arcuni, and S.R. Lundeen, Phys. Rev. A **38**, 135 (1988).

[13] W.G. Sturrus, E.A. Hessels, and S.R. Lundeen, Phys. Rev. Lett. **57**, 1863 (1986).

[14] E.A. Hessels, P.W. Arcuni, F.J. Deck, and S.R. Lundeen, Phys. Rev. A **46**, 2622 (1992).

[15] E.A. Hessels, F.J. Deck, P.W. Arcuni, and S.R. Lundeen, Phys. Rev. A **41**, 3663 (1990).

[16] P.W. Arcuni, Z.W. Fu, and S.R. Lundeen, Phys. Rev. A **42**, 6950 (1990).

[17] P.W. Arcuni, E.A. Hessels, and S.R. Lundeen, Phys. Rev. A **41**, 3648 (1990).

[18] L. Spruch, Phys. Today, 37 November (1986).

[19] J.F. Babb and L. Spruch, Phys. Rev. A **50**, 3848 (1994).

[20] A.Dalgarno, G.W.F. Drake, and G.A. Victor, Phys. Rev. **176**,194 (1968).

[21] M.J. Seaton and L. Steenman-Clark, J. Phys. B **10**,2639 (1977).

[22] R.J. Drachman, Phys. Rev. A **26**, 1228 (1982).

[23] W.G Schoenfeld, *Analysis of High-L Rydberg Levels of Silicon from the Solar Spectrum: A Test of the Extended Polarization Model*, PhD. thesis, Univ. of Mass. (1994), unpublished.

[24] Herzberg, G. and Ch. Jungen, J. Chem. Phys. **77**, 5876 (1982).

[25] Jungen, Ch., I. Dabrowski, G. Herzberg, and D. J. W. Kendall, 1989 J. Chem. Phys. **91**, 3926.

[26] Jungen, Ch., I. Dabrowski, G. Herzberg, and M. Vervloet, J. Chem. Phys. **93**, 2289 (1990).

[27] B. Zygelman, Phys. Rev. Lett. **64**, 256 (1990).

[28] R.E. Trees, Phys. Rev. **83** (1951), 756; **84**, 1089 (1951); **85**, 382 (1952).

[29] G. Racah, Phys. Rev. **85**, 381 (1952).

[30] W. Clark, C.H. Greene, and G. Miecznik, Phys. Rev. A **53**, 2248 (1996).

[31] W. Clark and C.H. Greene, Rev. Mod. Phys., (submitted 1998).

[32] B.J. Lyons, J.A. Shanchuck, J. Hostetler, and T.F. Gallagher, Phys. Rev. A **52**, 4586 (1995)

[33] W. Clark and C.H. Greene, Phys. Rev. A **56**, 403 (1997).

[34] W. Clark and C.H. Greene, Phys. Rev. A (in preparation).

[35] L. Pruvost, P. Camus, J.M. Lecomte, C.R. Mahon, and P. Pillet, J. Phys. B **24**, 4723 (1991).

[36] H. Lefebvre-Brion and R.W. Field, *Perturbations in the Spectra of Diatomic Molecules*, Academic Press, New York, 1986.

[37] H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, Mass. (1950).

[38] J.D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, New York (1975).

[39] H. Klar and U. Fano, Phys. Rev. Lett. **37**, 1132 (1976).

[40] H. Klar, Phys. Rev. A **15**, 1452 (1977).

[41] V.Aquilanti, S. Cavalli, and M.B. Sevryuk, J. Math. Phys. **35** (2), 536 (1994).

[42] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[43] R.N. Zare, *Angular Momentum.* John Wiley & Sons, Inc., New York, 1988.

[44] S. Watanabe and C.H. Greene, Phys. Rev. A **22**, 158 (1980).

[45] Gailitis, M. and R. Damburg, Proc. Phys. Soc. London **82**, 192 (1963).

[46] U. Fano and A.R.P. Rau, *Atomic Collisions and Spectra.* Academic Press, Inc., Orlando, Florida, 1986.

[47] H. Feshbach, Anals of Physics **5**, 357 (1958).

[48] H. Feshbach, Anals of Physics **19**, 287 (1962).

[49] L.S. Rodberg and R.M. Thaler, *Introduction to the Quantum Theory of Scattering.* Academic Press, Inc., New York, 1967.

[50] I. Lindgren and J. Morrison, *Atomic Many-Body Theory.* Springer Verlag, New York, 1982.

[51] D.A. Varshalovich, A.N. Moskalev, and V.K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988).

[52] A. Messiah, *Quantum Mechanics Volume II*, John Wiley & Sons, Inc., New York. 1966.

[53] U. Fano and J.H. Macek, Rev. Mod. Phys. **45**, 553 (1973).

[54] R.D. Cowan, *The Theory of Atomic Structure and Spectra*, University of California Press, Berkeley, 1981.

[55] C.H. Greene, *Fundamental Processes of Atomic Dynamics*, Plenum, New York, p105 (1988), C.H. Greene and L. Kim, Phys.Rev.A **38**, 5953 (1988), F. Robicheaux and C. H. Greene, Phys.Rev.A **46**, 3821 (1992)

[56] M. Aymar, C.H. Greene, and E. Luc-Koenig, Rev. Mod. Phys. **68**, 1015-1123 (1996).

[57] L.J. Kovalenko and J.B. Delos, J. Chem. Phys. **107**, 5460 (1997).

[58] L.J. Kovalenko and J.B. Delos, J. Chem. Phys. **107**, 5473 (1997).

[59] A. Hooker, C.H. Greene, and W. Clark, Phys. Rev. A **55**, 4609 (1997).

[60] M.V. Berry, Proc. Roy. Soc. London A **392**, 45 (1984).

[61] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).

[62] J. Moody, A. Shapere, and F. Wilczek, Phys. Rev. Lett. **56**, 893 (1986).

[63] R. Jackiw, Phys. Rev. Lett. **56**, 2779 (1986).

[64] R.G. Littlejohn and S. Weigert, Phys. Rev. A **48**, 924 (1993).

[65] A. Hibbert, M. Le Dourneuf, and Vo Ky Lan, J. Phys. B **10**, 1015 (1977).

[66] D. Sundholm and J. Olsen, Phys. Rev. A **49**, 3453 (1994).

[67] E.S. Chang, W.G. Schoenfeld, E. Biemont, P. Quinet, and P. Palmeri, Physica Scripta **49**, 26 (1994).

[68] C.Froese Fischer, Comp.Phys.Comm. **64**, 369 (1991)

[69] C.E. Moore, *Atomic Energy Levels*, Natl. Stand. Ref. Data. Ser., Nat. Bur. Stand. (U.S.), 35/Volume I, (1971).

[70] C.H. Greene and M. Aymar, Phys. Rev. A **44**, 1773 (1991).

[71] M. Aymar, private communication (1996).

[72] T.F. Gallagher, *Rydberg Atoms*. Cambridge University Press, New York, New York, 1994.

[73] W.C. Martin, and R. Zalubas, J. Phys. Chem. Ref. Data **9**, 4 (1980).

[74] C.E. Moore, *Atomic Energy Levels*, NBS Circular No. 467 (U.S. GPO, Washington, D.C., 1949), Vol. 3.

[75] M.D. Lindsay, C.-J. Dai, L.-T. Cai, T. F. Gallagher, F. Robicheaux, and C. H. Greene Phys. Rev. A **46**, 3789 (1992).

[76] R.T. Pack and J.O. Hirschfelder, J. Chem. Phys. **49**, 4009 (1968).

[77] B. R. Judd *Angular Momentum Theory for Diatomic Molecules*, Academic Press, Inc., New York, 1975.

[78] A. Carrington and R.A. Kennedy, *Gas Phase Ion Chemistry Vol. 3 Ions and Light*, Academic Press, Inc., London (1984).

[79] A. Carrington, I.R. McNAb, and C.A. Montgomerie, J. Phys. B **22**, 3551 (1989).

[80] G. Hunter, B.F. Gray, and H.O. Pritchard, J. Chem. Phys. **45**, 3806 (1966).

[81] G. Hunter and H.O. Pritchard, J. Chem. Phys. **46**, 2146 (1966).

[82] G. Hunter and H.O. Pritchard, J. Chem. Phys. **46**, 2153 (1967).

[83] P.B. Davies, M.A. Guest, and R.J. Stickland, J. Chem. Phys. **93**, 5408 (1990).

[84] P.B. Davies, M.A. Guest, and R.J. Stickland, J. Chem. Phys. **93**, 5417 (1990).

[85] D.M. Bishop and B. Lam, Mol. Phys. **65**, 679 (1988).

[86] I.I. Sobelman, *Introduction to the Theory of Atomic Spectra*, Pergamon Press, New York, 1972.

[87] F. Friedrich, *Theoretical Atomic Physics*, Springer-Verlag, New York, 1990.

[88] H.A. Bethe and E.E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, Academic Press, New York, 1957.

[89] R.J. Drachman, Phys. Rev. A **33**, 2780 (1986).

[90] R.J. Drachman, Phys. Rev. A **37**, 979 (1988).

[91] D.M. Bishop and Lap. M. Cheung, Phys. Rev. A **16**, 640 (1977).

[92] D.M. Bishop and R.W. Wetmore, Mol. Phys. **26**, 145 (1973).

[93] R.E. Moss, Chem. Phys. Lett. **172**, 458 (1990).

[94] W. Kolos and L. Wolniewicz, Rev. Mod. Phys. **35**, 473 (1963).

[95] W. Kolos and L. Wolniewicz, J. Chem. Phys. **41**, 3663 (1964).

[96] W. Kolos and L. Wolniewicz, J. Chem. Phys. **49**, 404 (1968).

[97] L. Wolniewicz and J.D. Poll, J. Chem. Phys. **73**, 6225 (1980).

[98] D.M. Bishop and L.M. Cheung, J. Phys. B **11**, 31333 (1978).

[99] D.M. Bishop and Lap. M. Cheung, J. Phys. B **12**, 3135 (1979).

[100] D.M. Bishop, L.M. Cheung, and A.D. Buckingham, Mol. Phys. **41**, 1225 (1980).

[101] D.M. Bishop and B Lam, Chem. Phys. Lett. **134**, 283 (1987).

[102] D.M. Bishop, Phys. Rev. Lett. **62**, 3008 (1989).

[103] J. Shertzer and J. DeGuzman, (in preparation).

[104] W.G. Sturrus, P.L. Jacobson, and S.R. Lundeen, (in preparation).

[105] R.P. McEachran, C.J. Veenstra, and M. Cohen, Chem. Phys. Lett. **59**, 275 (1978).

[106] E.S. Chang, S. Pulchtopek, and E.E. Eyler, J. Chem. Phys. **80**, 601 (1984).

[107] D.M. Bishop, B. Kirtman, B. Champagne, J. Chem. Phys. **107**, 5780 (1997).

[108] R.J. Drachman, Phys. Rev. A **31**, 1253 (1985).

[109] M. Brieger, Chem. Phys. **89**, 275 (1989).

[110] C.W. Clark and K.T. Taylor, J. Phys. B **15**, 1175 (1982).

[111] G.B. Arfken and H.J. Weber, *Mathematical Methods For Physicists, Forth Edition*, Academic Press, New York, 1995.

[112] M.L. Boas, *Mathematical Methods In The Physical Sciences, Second Edition*, John Wiley & Sons, New York, 1983.

[113] J. Mathews and R.L. Walker, *Mathematical Methods of Physics*, Addison-Wesley Publishing Company, Inc., New York, 1970.

APPENDIX A

PARAMETERS FOR HIGH-ℓ RYDBERG STATES OF LOW-Z ATOMS

Here we give explicit expressions for all radial parameters used in Eqs.(6.1) and (6.2). The constant versions of these radial parameters, those used in Eqs.(4.40) and (4.41), are obtained by taking the limit $r \rightarrow \infty$, or by replacing $r_>$ with the radial coordinate r of the Rydberg electron. The summations over i and j in each expression refer to summations over all core electrons. $r_< = min\{r_i, r\}$ and $r_> = max\{r_i, r\}$, where r refers to the Rydberg electron. The summations over γ_ν denote summations over all bound, and integrals over all continuum states of the core that are not degenerate with E_{μ_0}.

The net Coulomb charge seen by the Rydberg electron as a function of the radial distance is

$$C_c(r) = r\langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{1}{r_>} C^{(0)}(\hat{r}_i) \| \gamma_c L_c \rangle - Z. \tag{A.1}$$

The quadrupole moment of the ionic core is

$$Q(r) = -r^3 \begin{pmatrix} L_c & 2 & L_c \\ -L_c & 0 & L_c \end{pmatrix} \langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{r_<^2}{r_>^3} C^{(2)}(\hat{r}_i) \| \gamma_c L_c \rangle. \tag{A.2}$$

The scalar dipole polarizability of the ionic core is

$$\alpha_s(r) = \frac{-2}{3(2L_c+1)} r^4 \sum_{\gamma_\nu, L_\nu} \frac{(-1)^{L_c+L_\nu}}{E_{\mu_0} - E_\nu} \langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_c} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \tag{A.3}$$

The second-rank tensor dipole polarizability of the ionic core is

$$\alpha_t(r) = -2\sqrt{\frac{10}{3}}r^4 \begin{pmatrix} L_c & 2 & L_c \\ -L_c & 0 & L_c \end{pmatrix} \sum_{\gamma_\nu, L_\nu} \frac{1}{E_{\mu_0} - E_\nu} \begin{Bmatrix} 2 & L_c & L_c \\ L_\nu & 1 & 1 \end{Bmatrix}$$

$$\langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle \langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_c} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \tag{A.4}$$

The scalar hyperpolarizability is

$$\eta_s(r) = \frac{-1}{5(2L_c+1)} \left[\bar{\eta}_s^I(r) + \bar{\eta}_s^{II}(r) \right], \tag{A.5}$$

where

$$\eta_s^I(r) = r^6 \sum_{\gamma_\nu, L_\nu} \frac{(-1)^{L_c+L_\nu}}{E_{\mu_0} - E_\nu} \langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{r_<^2}{r_>^3} C^{(2)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_c} \frac{r_<^2}{r_>^3} C^{(2)}(\hat{r}_j) \| \gamma_c L_c \rangle, \tag{A.6}$$

and

$$\eta_s^{II}(r) = -\frac{5}{6} r^6 \sum_{\gamma_\nu, L_\nu} \frac{(-1)^{L_c+L_\nu}}{(E_{\mu_0} - E_\nu)^2} \langle \gamma_c L_c \| \sum_{i=1}^{N_c} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_c} \left(\frac{d^2}{dr^2} \frac{r_<}{r_>^2} \right) C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \tag{A.7}$$

The tensor hyperpolarizability is

$$\eta_t(r) = \begin{pmatrix} L_c & 2 & L_c \\ -L_c & 0 & L_c \end{pmatrix} \left[\bar{\eta}_t^I(r) + \bar{\eta}_t^{II}(r) + \bar{\eta}_t^{III} \right], \tag{A.8}$$

where

$$\eta_t^I(r) = \sqrt{\frac{10}{7}} r^6 \sum_{\gamma_\nu, L_\nu} \frac{1}{E_{\mu_0} - E_\nu} \begin{Bmatrix} 2 & L_c & L_c \\ L_\nu & 2 & 2 \end{Bmatrix} \langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<^2}{r_>^3} C^{(2)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \frac{r_<^2}{r_>^3} C^{(2)}(\hat{r}_j) \| \gamma_c L_c \rangle, \tag{A.9}$$

$$\eta_t^{II}(r) = 2\sqrt{\frac{15}{7}} r^6 \sum_{\gamma_\nu, L_\nu} \frac{1}{E_{\mu_0} - E_\nu} \begin{Bmatrix} 2 & L_c & L_c \\ L_\nu & 1 & 3 \end{Bmatrix} \langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \frac{r_<^3}{r_>^4} C^{(3)}(\hat{r}_j) \| \gamma_c L_c \rangle, \tag{A.10}$$

and

$$\eta_t^{III}(r) = \sqrt{\frac{5}{6}} r^6 \sum_{\gamma_\nu, L_\nu} \frac{1}{(E_{\mu_0} - E_\nu)^2} \begin{Bmatrix} 2 & L_c & L_c \\ L_\nu & 1 & 1 \end{Bmatrix} \langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \left(\frac{d^2}{dr^2} \frac{r_<}{r_>^2} \right) C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \tag{A.11}$$

The scalar nonadiabatic dipole polarizability is

$$\beta_s(r) = \frac{r^4}{6(2L_c+1)} \sum_{\gamma_\nu, L_\nu} \frac{(-1)^{L_c+L_\nu}}{(E_{\mu_0} - E_\nu)^2} \langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle$$

$$\langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \tag{A.12}$$

The tensor nonadiabatic dipole polarizability is

$$\beta_t(r) = \sqrt{\frac{5}{6}} r^4 \begin{pmatrix} L_c & 2 & L_c \\ -L_c & 0 & L_c \end{pmatrix} \sum_{\gamma_\nu, L_\nu} \frac{1}{(E_{\mu_0} - E_\nu)^2} \begin{Bmatrix} 2 & L_c & L_c \\ L_\nu & 1 & 1 \end{Bmatrix}$$

$$\langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle \langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \qquad \text{(A.13)}$$

The vector dipole hyperpolarizability is

$$\beta_v(r) = \sqrt{\frac{3}{2}} \frac{r^4}{\sqrt{L_c(L_c+1)(2L_c+1)}} \sum_{\gamma_\nu, L_\nu} \frac{1}{(E_{\mu_0} - E_\nu)^2} \begin{Bmatrix} 1 & L_c & L_c \\ L_\nu & 1 & 1 \end{Bmatrix}$$

$$\langle \gamma_c L_c \| \sum_{i=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_i) \| \gamma_\nu L_\nu \rangle \langle \gamma_\nu L_\nu \| \sum_{j=1}^{N_e} \frac{r_<}{r_>^2} C^{(1)}(\hat{r}_j) \| \gamma_c L_c \rangle. \qquad \text{(A.14)}$$

The angular factors are given by

$$\mathcal{A}^{(2)}_{\mu\mu'} = [(2\ell+1)(2\ell'+1)(2J_c+1)(2J_c'+1)]^{\frac{1}{2}}$$

$$(-1)^{J_c'+L_c+S_c} \begin{pmatrix} \ell' & \ell & 2 \\ 0 & 0 & 0 \end{pmatrix} \Big/ \begin{pmatrix} L_c & 2 & L_c \\ -L_c & 0 & L_c \end{pmatrix}$$

$$(-1)^{J_c'+K} \begin{Bmatrix} J_c & \ell & K \\ \ell' & J_c' & 2 \end{Bmatrix} \begin{Bmatrix} 2 & L_c & L_c \\ S_c & J_c & J_c' \end{Bmatrix}, \qquad \text{(A.15)}$$

and

$$\mathcal{B}^{(0)}_{\mu\mu'} = (\ell(\ell+1)+2)\,\delta_{\mu\mu'}. \qquad \text{(A.16)}$$

and, finally,

$$\mathcal{B}^{(2)}_{\mu\mu'} = \mathcal{A}^{(2)}_{\mu\mu'} (-1)^{\ell+\ell'} \Big/ \left[\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell' & \ell & 2 \\ 0 & 0 & 0 \end{pmatrix} \right]$$

$$\sum_{\ell_\nu} (-1)^{\ell_\nu} \ell_\nu (\ell_\nu+1)(2\ell_\nu+1) \begin{pmatrix} \ell & 1 & \ell_\nu \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_\nu & 1 & \ell' \\ 0 & 0 & 0 \end{pmatrix} \begin{Bmatrix} 1 & 1 & 2 \\ \ell' & \ell & \ell_\nu \end{Bmatrix} \qquad \text{(A.17)}$$

APPENDIX B

CHANNEL PARAMETERS FOR RYDBERG STATES OF H_2 AND D_2

In this appendix we give explicit expressions for all permanent multipole moments and induced polarizabilities and hyperpolarizabilities used in Eqs.(7.18) and (7.19). The summations over ν denote summations over all bound and continuum states Ψ_ν of the molecular ions that are not degenerate with E_{μ_0}. These expressions are formulated specifically for molecular ion wavefunctions computed using the adiabatic nuclei approximation.

The k-th permanent multipole moment of the molecular ion is

$$Q^{(k)}_{mu\mu'} = \langle \Psi_\mu \mid \left(M_k R^k - N_k r_e C^{(k)}_0(\hat{r}_e) \right) \mid \Psi_{\mu'} \rangle \tag{B.1}$$

where the nuclear mass term M_k and the electronic mass term N_k are defined in Chapter 7. The scalar dipole polarizability of the molecular ion is

$$\alpha_{s_{\mu\mu}} = -\frac{2}{3} N_1^2 \sum_\nu \frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu} \begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}^2 \mid \langle \Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu \rangle \mid^2 . \tag{B.2}$$

The second-rank tensor dipole polarizability of the molecular ion is

$$\alpha_{t_{\mu\mu'}} = -10\sqrt{\frac{2}{15}} N_1^2 \Big/ \begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+} \frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu} \begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix} \begin{pmatrix} N_\nu^+ & 1 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 1 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix} \langle \Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu \rangle \langle \Psi_\nu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_{\mu'} \rangle^*$$

(B.3)

The scalar hyperpolarizability is

$$\eta_{s_{\mu\mu}} = -\frac{N_2^2}{5}\sum_{\nu}\frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu}\begin{pmatrix} N^+ & 2 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}^2 |\langle\Psi_\mu | r_e^2 C^{(2)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_\nu\rangle|^2$$

$$+N_1^2\sum_{\nu}\frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}^2 |\langle\Psi_\mu | r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_\nu\rangle|^2 \quad \text{(B.4)}$$

The second-rank tensor hyperpolarizability is

$$\eta_{t_{\mu\mu'}} = 5\sqrt{\frac{2}{35}}N_2^2 \Big/ \begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{\nu}(-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu}\begin{pmatrix} N^+ & 2 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 2 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 2 & 2 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu | r_e^2 C^{(2)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_\nu\rangle\langle\Psi_\nu | r_e^2 C^{(2)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_{\mu'}\rangle^*$$

$$+5\sqrt{\frac{3}{35}}N_1 N_3 \Big/ \begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{\nu}(-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 3 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 3 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu | r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_\nu\rangle\langle\Psi_\nu | r_e^3 C^{(3)}_{-\lambda_\nu}(\hat{r}_e) | \Psi_{\mu'}\rangle^*$$

$$+5\sqrt{\frac{3}{35}}N_3N_1 \Big/ \begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+} \frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu} \begin{pmatrix} N^+ & 3 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix} \begin{pmatrix} N_\nu^+ & 1 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 3 & 1 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix} \langle \Psi_\mu \mid r_e^3 C_{-\lambda_\nu}^{(3)}(\hat{r}_e) \mid \Psi_\nu \rangle \langle \Psi_\nu \mid r_e C_{-\lambda_\nu}^{(1)}(\hat{r}_e) \mid \Psi_{\mu'} \rangle^*$$

$$-15\sqrt{\frac{2}{15}}N_1^2 \Big/ \begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+} \frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2} \begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix} \begin{pmatrix} N_\nu^+ & 1 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 1 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix} \langle \Psi_\mu \mid r_e C_{-\lambda_\nu}^{(1)}(\hat{r}_e) \mid \Psi_\nu \rangle \langle \Psi_\nu \mid r_e C_{-\lambda_\nu}^{(1)}(\hat{r}_e) \mid \Psi_{\mu'} \rangle^* \tag{B.5}$$

The fourth-rank tensor hyperpolarizability is

$$\lambda_{h_{\mu\mu'}} = -9\sqrt{\frac{2}{35}}N_2^2 \Big/ \begin{pmatrix} N^+ & 4 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+} \frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu} \begin{pmatrix} N^+ & 2 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix} \begin{pmatrix} N_\nu^+ & 2 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 2 & 2 & 4 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix} \langle \Psi_\mu \mid r_e^2 C_{-\lambda_\nu}^{(2)}(\hat{r}_e) \mid \Psi_\nu \rangle \langle \Psi_\nu \mid r_e^2 C_{-\lambda_\nu}^{(2)}(\hat{r}_e) \mid \Psi_{\mu'} \rangle^*$$

$$-\frac{6}{\sqrt{7}}N_1N_3\Bigg/\begin{pmatrix} N^+ & 4 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 3 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 3 & 4 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu\rangle\langle\Psi_\nu \mid r_e^3 C^{(3)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_{\mu'}\rangle^*$$

$$-\frac{6}{\sqrt{7}}N_3N_1\Bigg/\begin{pmatrix} N^+ & 4 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_\nu (-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{E_{\mu_0} - E_\nu}\begin{pmatrix} N^+ & 3 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 1 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 3 & 1 & 4 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu \mid r_e^3 C^{(3)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu\rangle\langle\Psi_\nu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_{\mu'}\rangle^* \tag{B.6}$$

The scalar nonadiabatic dipole polarizability is

$$\beta_{s_{\mu\mu}} = \frac{N_1^2}{6}\sum_\nu \frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}^2 \mid \langle\Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu\rangle \mid^2 . \tag{B.7}$$

The second-rank nonadiabatic dipole polarizability is

$$\beta_{t_{\mu\mu'}} = -\sqrt{\frac{5}{3}}N_1^2\Bigg/\begin{pmatrix} N^+ & 2 & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{\nu}(-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 1 & N^{+\prime} \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 1 & 2 \\ N^{+\prime} & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu\rangle\langle\Psi_\nu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_{\mu'}\rangle^*. \tag{B.8}$$

The vector hyperpolarizability of the molecular ion is

$$\beta_{v_{\mu\mu}} = (-1)^{N^+} N_1^2\sqrt{\frac{3}{2}}\left[\frac{(2N^+ + 1)}{N^+(N^+ + 1)}\right]^{\frac{1}{2}}$$

$$\sum_{\nu}(-1)^{N_\nu^+}\frac{(2N_\nu^+ + 1)}{(E_{\mu_0} - E_\nu)^2}\begin{pmatrix} N^+ & 1 & N_\nu^+ \\ 0 & -\lambda_\nu & \lambda_\nu \end{pmatrix}\begin{pmatrix} N_\nu^+ & 1 & N^+ \\ -\lambda_\nu & \lambda_\nu & 0 \end{pmatrix}$$

$$\begin{Bmatrix} 1 & 1 & 1 \\ N^+ & N^+ & N_\nu^+ \end{Bmatrix}\langle\Psi_\mu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\nu\rangle\langle\Psi_\nu \mid r_e C^{(1)}_{-\lambda_\nu}(\hat{r}_e) \mid \Psi_\mu\rangle^* \tag{B.9}$$

The angular coupling factors are

$$\mathcal{P}^k_{\mu\mu'} = \langle(N^+\ell)NM_N \mid C^{(k)}(\hat{R})\cdot C^{(k)}(\hat{r}) \mid (N^{+\prime}\ell')N'M_{N'}\rangle$$

$$= \delta_{NN'}\delta_{M_N M'_N}(-1)^{N^+ + N^{+\prime} N}\left[(2N^+ + 1)(2N^{+\prime} + 1)(2\ell + 1)(2\ell' + 1)\right]^{\frac{1}{2}}$$

$$\begin{pmatrix} N^+ & k & N^{+\prime} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \ell & k & \ell' \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} N^+ & \ell & N \\ \ell' & N^{+\prime} & k \end{Bmatrix}, \tag{B.10}$$

and

$$\mathcal{B}^{(0)}_{\mu\mu'} = [\ell(\ell + 1) + 2]\,\delta_{\mu\mu'}, \tag{B.11}$$

and

$$B^{(2)}_{\mu\mu'} = P^{k}_{\mu\mu'}(-1)^{\ell+\ell'}\sum_{\ell_\nu}(-1)^{\ell_\nu}\ell_\nu(\ell_\nu+1)(2\ell_\nu+1)$$

$$\begin{pmatrix}\ell & 1 & \ell'\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell & 1 & \ell'\\ 0 & 0 & 0\end{pmatrix}\begin{Bmatrix}1 & 1 & 2\\ \ell' & \ell & \ell_\nu\end{Bmatrix}$$

$$\Big/\left[\begin{pmatrix}1 & 1 & 2\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell & 2 & \ell'\\ 0 & 0 & 0\end{pmatrix}\right] \qquad \text{(B.12)}$$

and finally

$$\langle(N^+\ell)NM_N \mid \vec{N}^+\cdot\vec{\ell}\mid (N^{+\prime}\ell')N'M_{N'}\rangle = \delta_{NN'}\delta_{M_N M'_N}\delta_{N+N^{+\prime}}\delta_{\ell\ell'}(-1)^{N+\ell+N^+}$$

$$\left[N^+(N^++1)(2N^++1)\ell(\ell+1)(2\ell+1)\right]^{\frac{1}{2}}\begin{Bmatrix}N^+ & \ell & N\\ \ell' & N^{+\prime} & 1\end{Bmatrix}. \qquad \text{(B.13)}$$

APPENDIX C

STURMIAN BASIS

The Sturmian Basis is extremely useful in almost any problem involving Coulomb-like potentials. The Sturmian functions $S_{n\ell}^{(k)}(r)$ satisfy the ordinary differential equation [110]

$$\left[-\frac{1}{2}\frac{d^2}{dr^2}+\frac{\ell(\ell+1)}{2r^2}+\frac{k^2}{2}\right]S_{n\ell}^{(k)}(r)=\frac{nk}{r}S_{n\ell}^{(k)}(r) \tag{C.1}$$

where k is a constant that can be chosen to adjust the spatial extent of the Sturmian functions. Provided k is not n-dependent the Sturmians form a $\mathcal{L}^2$ representation.

The Sturmian functions are given explicitly by

$$S_{n\ell}^{(k)}(r)=N(2kr)^{\ell+1}e^{-kr}L_{n-\ell-1}^{(2\ell+1)}(2kr) \tag{C.2}$$

where the $L_{n-\ell-1}^{(2\ell+1)}(2kr)$ are Laguerre functions. To determine the normalization constant N and establish the completeness of the Sturmian functions it is useful to state some of the properties of Laguerre functions. The normalization of the Laguerre functions $\{L_n^{(\alpha)}(x)\}$ [111, 112] is given by

$$\int_0^\infty dx e^{-x}x^{\alpha+1}L_n^{(\alpha)}(x)L_n^{(\alpha)}(x)=\frac{(n+\alpha)!}{n!}(2n+\alpha+1) \tag{C.3}$$

and their orthogonality is given by

$$\int_0^\infty dx e^{-x}x^{\alpha}L_n^{(\alpha)}(x)L_m^{(\alpha)}(x)=\frac{(n+\alpha)!}{n!}\delta_{nm}. \tag{C.4}$$

From the normalization condition for the Laguerre functions it is easy to show that the Sturmian normalization constant is

$$N = \left[\frac{k(n-\ell-1)!}{n(n+\ell)!}\right]^{\frac{1}{2}} \tag{C.5}$$

so that

$$\int_0^\infty dr S_{n\ell}^{(k)}(r) S_{n\ell}^{(k)}(r) = 1. \tag{C.6}$$

In order to show that the Sturmian functions form a complete orthonormal set, as defined by Sturm-Liousville Theory [111, 113], we need to find a weight function such that

$$\int_0^\infty dr \rho(r) S_{n\ell}^{(k)}(r) S_{n\ell}^{(k)}(r) = 0. \tag{C.7}$$

Using the orthogonality condition for Laguerre functions it is easy to show that the Sturmian weight function is given by

$$\rho(r) = \frac{n}{kr}. \tag{C.8}$$

The the formal completeness relation for the Sturmian functions is

$$\sum_{n=\ell+1}^{\infty} \frac{n}{k(rr')^{\frac{1}{2}}} S_{n\ell}^{(k)}(r) S_{n\ell}^{(k)}(r') = \delta(r-r'). \tag{C.9}$$

APPENDIX D

PROLATE SPHEROIDAL COORDINATES

The prolate spheroidal coordinates are particularly useful in the description of two center problems [77, 111]. Let the two centers be located at distances r_a and r_b from a common origin. The orthogonal prolate spheroidal coordinates are defined in terms of a radial coordinates ξ

$$\xi \equiv \frac{r_a + r_b}{R} \quad \text{with} \quad 1 \leq \xi < \infty, \tag{D.1}$$

an angular coordinate η

$$\eta \equiv \frac{r_a - r_b}{R} \quad \text{with} \quad -1 \leq \eta \leq 1, \tag{D.2}$$

and an azimuthal coordinate ϕ

$$\phi \quad \text{with} \quad 0 \leq \phi \leq 2\pi. \tag{D.3}$$

The differential volume elements is

$$d\tau = \frac{R^3(\xi^2 - \eta^2)}{8} d\xi d\eta d\phi. \tag{D.4}$$

The Laplacian is

$$\nabla^2 = \frac{4}{R^2(\xi^2 - \eta^2)} \left\{ \frac{\partial}{\partial \xi}(\xi^2 - 1)\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}(1 - \eta^2)\frac{\partial}{\partial \eta} + \frac{(\xi^2 - \eta^2)}{(\xi^2 - 1)(1 - \eta^2)}\frac{\partial^2}{\partial \phi^2} \right\}. \tag{D.5}$$

The use of prolate spheroidal coordinates in the generalized hydrogen molecular ion facilitates the separation of the electronic Schrödinger equation into three equations coupled by separation constants.

APPENDIX E

FINITE ELEMENT METHOD

The method of finite elements belongs to a class of numerical techniques used to solve ordinary and partial differential equations. In general, the method transforms a set of coupled differential equations into a variational, generalized eigenvalue problem. Depending on how the problem is formulated, a discrete set of bound and continuum energy eigenstates can be obtained, or by chosing the eigenvalue to be the log-derivative of the wavefunction at the boundary of the system, continuous energy eigenstates can be obtained. Here the finite element method for a one-dimensional Schrödinger equation with a restricted configuration space is developed, corresponding to the first case where discrete bound and continuum states are desired.

The Schrödinger equation for an arbitray potential $V(r)$, for $0 \leq r < \infty$, is, in atomic units,

$$\left(-\frac{1}{2}\frac{d^2}{dr^2} + V(r) - E\right)\psi(r) = 0. \tag{E.1}$$

Let us define an integral S

$$S \equiv \int_0^\infty dr\psi(r)\left(-\frac{1}{2}\frac{d^2}{dr^2} + V(r) - E\right)\psi(r). \tag{E.2}$$

Integrating by parts this becomes

$$S = \int_0^\infty dr\left[\psi(r)\left(V(r) - E\right)\psi(r) + \frac{1}{2}\left(\frac{d\psi(r)}{dr}\right)^2\right] + \left[\frac{1}{2}\psi(r)\frac{d\psi(r)}{dr}\right]_0^\infty. \tag{E.3}$$

Now for most problems of interest $lim_{r\to 0}\psi(r) = 0$. However, let us impose the condition that $lim_{r\to\infty}\frac{d\psi(r)}{dr} = 0$ so that the surface term, obtained from the integration by parts, vanishes. This final boundary condition is appropriate for bound states, but discretizes all continuum states (Pseudo States).

Let us now define an action A

$$A \equiv \frac{1}{2}\left(\frac{d\psi(r)}{dr}\right)^2 + \psi(r)\left(V(r) - E\right)\psi(r). \tag{E.4}$$

Then by the Euler-Lagrange equation

$$\frac{dA}{d\psi} - \frac{d}{dr}\frac{dA}{d\left(\frac{d\psi(r)}{dr}\right)} = 2\left(V(r) - E\right)\psi(r) - \frac{d^2}{dr^2}\psi(r) = 0 \tag{E.5}$$

vanishes since the last expression is just the original Schrödinger equation. Thus $\delta S = 0$ and S is a variational equation.

The power of the finite element method is derived by chosing a basis set that provides both computational accuracy and efficiency. Imagine that the interval $0 \leq r < \infty$ is broken into M sectors and that in each sector we expand the wavefunction $\psi(r)$ in a basis. Let us choose the basis to be six polynomials $\{f_i(r)\}$ with the constraints that one is nonzero at any boundary and another has a nonzero derivative at any boundary in a given sector as show in Figure E.1. That is, let us expand the wavefunction as

$$\psi(r) = \sum_{m=1}^{M}\sum_{i=1}^{6} c_i^m f_i^{(m)}(r), \tag{E.6}$$

where the $\{f_i^{(m)}(r)\}$ is a set of six polynomials defined within the m-th sector boundaries. Since $\delta S = 0$

$$\frac{\partial S}{\partial c_i^m} = 0, \tag{E.7}$$

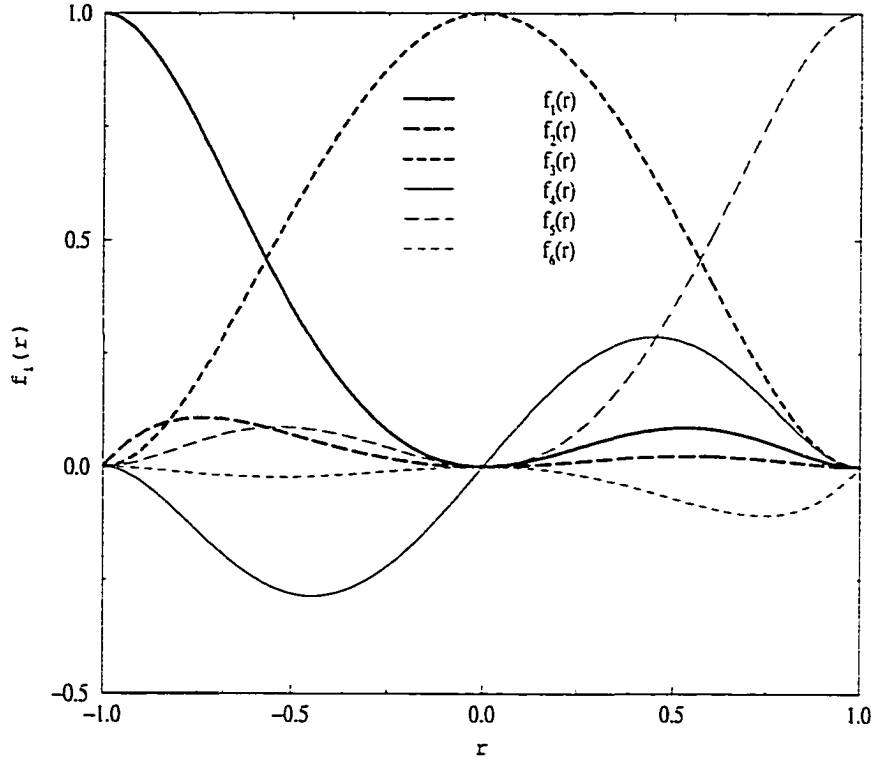


Figure E.1: Polynomial basis set used in finite element method.

where the $\{c_i^m\}$ are interpreted as variational parameters. Substituting the $\psi(r)$ expansion, Eq.[E.6], into the expression for S, Eq.[E.2], and using this partial derivative expression we find the generalized eigenvalue equation

$$\sum_{m=1}^{M}\sum_{j=1}^{6}\int_{r_{m-1}}^{r_m} dr^{(m)}\left[\frac{1}{2}\frac{df_i^{(m)}(r)}{dr}\frac{df_j^{(m)}(r)}{dr}+f_i^{(m)}(r)V(r)f_j^{(m)}(r)\right]c_j^m$$

$$= E\sum_{m=1}^{M}\sum_{j=1}^{6}\int_{r_{m-1}}^{r_m} dr f_i^{(m)}(r)f_j^{(m)}(r)c_j^m. \qquad \text{(E.8)}$$

Since we have chosen to partition the configurations space, in addition to expanding the wavefunction in a discrete basis, the wavefunction and its derivative must be forced to be continuous across each sector boundary. As a result, not all of the c_i^m coefficients are independent. However, the nature of the polynomial basis set ensures that in any sector only the first two coefficients are coupled to the previous sector, and the last two coefficients are coupled to the next sector.

The wavefunction and its derivative in the m th sector are

$$\psi^{(m)}(r) = \sum_{i=1}^{6} c_i^m f_i^{(m)}(r), \qquad \text{(E.9)}$$

$$\frac{d}{dr}\psi^{(m)}(r) = \sum_{i=1}^{6} c_i^m \frac{df_i^{(m)}}{dr}(r). \qquad \text{(E.10)}$$

Continuity of the wave function between sectors m and $m+1$, at the boundary coordinate r_m, is given by

$$\sum_{i=1}^{6} c_i^m f_i^{(m)}(r_m) = \sum_{j=1}^{6} c_j^{m+1} f_j^{(m+1)}(r_m) \qquad \text{(E.11)}$$

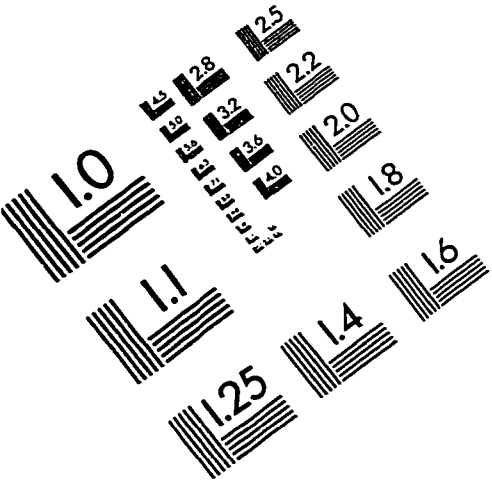
and the continuity of the derivative of the wavefunction across the boundary is

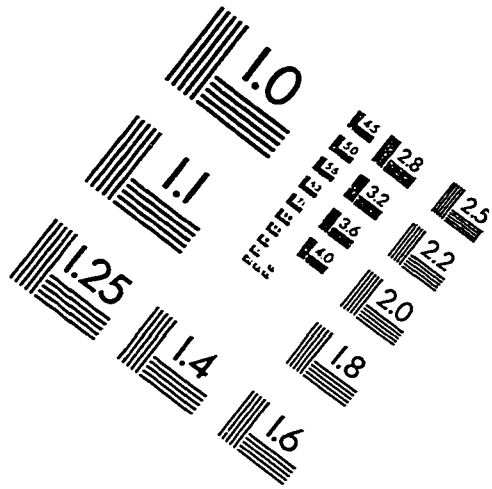
$$\sum_{i=1}^{6} c_i^m \frac{df_i^{(m)}}{dr}(r_m) = \sum_{j=1}^{6} c_j^{m+1} \frac{df_j^{(m+1)}}{dr}(r_m). \tag{E.12}$$

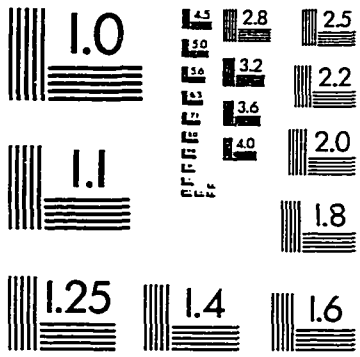
From these equations and the choice of our polynomial basis, the relationship among the sector coefficients is

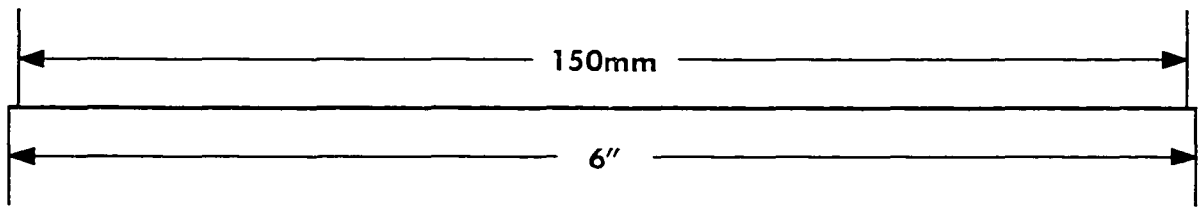
$$c_5^m = c_1^{m+1} \quad \text{and} \quad c_6^m = c_2^{m+1}. \tag{E.13}$$

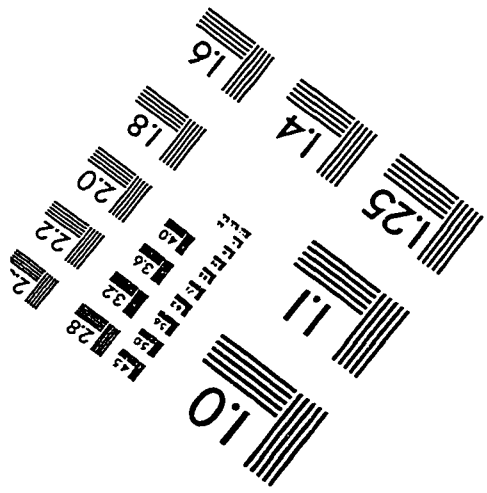
IMAGE EVALUATION
TEST TARGET (QA-3)

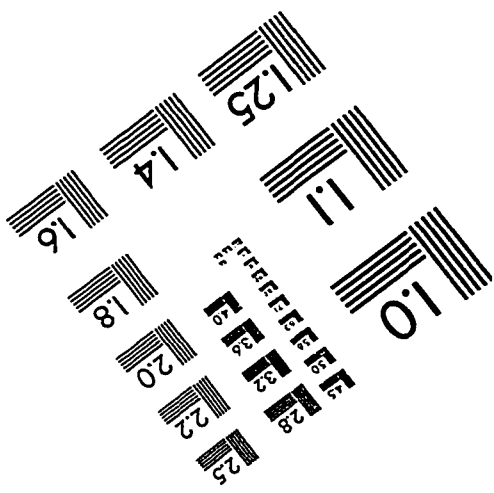












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Io's interaction with the jovian magnetosphere: Models of particle acceleration and scattering

by

Frank Judson Crary

B. A., University of California, Berkeley, 1991

M. S., University of Colorado, Boulder, 1994

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Astrophysical and Planetary Science

1998

This thesis entitled:
Io's interaction with the jovian magnetosphere: Models of particle acceleration and scattering
written by Frank Judson Crary
has been approved for the Department of Astrophysical and Planetary Science

Fran Bagenal

Nicholas M. Schneider

Date April 23, 1998

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Crary, Frank Judson (Ph. D., Astrophysical, Planetary and Atmospheric Science)

Io's interaction with the jovian magnetosphere: Models of particle acceleration and scattering

Thesis directed by Professor Fran Bagenal

I develop models of electron acceleration and ion scattering which result from Io's interaction with the jovian magnetosphere. According to my models, Io initially generates transient currents and an Alfvénic disturbance when it first encounters a jovian magnetic field line, and the interaction would eventually settle into a system of steady Birkeland currents as the field line is advected downstream past Io and into Io's wake. I derive a model of wave propagation and electron acceleration by the Alfvénic transient, due to electron inertial effects. My numerical calculations show that the power and particle energy of the resulting electron beam are consistent with observations of the Io-related auroral spot and of Jupiter's S-burst decametric emissions. In the case of the steady currents and Io's wake, I show that these currents would drive instabilities and argue that electrostatic double layers would form in the high latitudes of the Io/Io wake flux tubes. I examine the role of these double layers in producing energetic electrons and estimate the likely electron energies and power. This model agrees with observations of a long arc in the jovian aurora, extending away from the Io-related spot, the L-burst decametric radio emissions and electron beams observed by the Galileo spacecraft in Io's wake. Finally, I consider the Galileo observations of ion cyclotron waves near Io. I use the absence of waves near the S and O gyrofrequencies to place limits on the source rate of heavy ions near Io. For a sufficiently low source rate, the thermal core population prevents ion cyclotron instabilities and wave growth. I use these limits to constrain the neutral column density of Io's exosphere and amount of plasma produced within 2 to 10 body radii of Io.

Dedication

This thesis is dedicated to C. K. Goertz and E. M. Shoemaker

Acknowledgements

While working on this thesis, I have benefited from conversations and collaborations with a wide range of people. Here at the University of Colorado, this includes my advisor, Fran Bagenal, as well as David Brain, Matt Burger, Mihaly Horanyi, Nick Schneider Martin Taylor, and Jody Wilson. On the subject of the Io interaction, Jon Linker and Fritz Neubauer; on plasma re-acceleration and ionospheric coupling, Tom Hill and Duane Pontius; and on the subject of ion cyclotron wave scattering, Debbie Huddleston and Jörge Warnecke.

Contents

Chapter

1 Introduction 1

1.1 Overview 2

1.2 The character of Io's electrodynamic interaction 5

1.3 Io-related radio emissions 8

1.4 Io's auroral spot 11

1.5 Ion cyclotron waves near Io 17

1.6 Plasma conditions along Io's flux tube 19

1.7 Overview of thesis 28

2 Electron acceleration by Alfvén waves 30

2.1 Introduction 31

2.2 Estimates of the Alfvénic Disturbance 32

2.3 Current-Carrying Alfvén Waves 34

2.4 Propagation in a Nonuniform Medium 38

2.5 Acceleration of an Electron Beam 48

2.6 Approximate Electron Flux and Energy 51

2.7 Comparison to Observations 59

2.8 Summary 61

3 Particle acceleration by a steady current system 64

3.1 Introduction 65

3.2 Current closure within an ionosphere 66

3.3 Double Layers 70

3.4 Double Layers along the Io flux tubes 80
3.5 Double layers along the Io wake flux tubes 85
3.6 Summary 89

4 Ion cyclotron waves near Io 91
4.1 Introduction 92
4.2 Ion scattering and Quasi-linear diffusion 98
4.3 Limits on non-thermal S^+ and O^+ abundance 103
4.4 Test particle scattering analysis 104
4.5 Io's neutral exosphere 121
4.6 Summary 128

5 Summary 131
5.1 Results 132
5.2 Future modeling 134
5.3 Future observations 136

Bibliography 141

Figures

Figure

1.1 Overview of the Io's interaction with the jovian magnetosphere. Plasma density is shown in purple, and based on the Bagenal (1994) model. HST/STIS images of the jovian aurora and Io auroral spot and Galileo SSI image of Io. 3

1.2 A sketch of the Io interaction's geometry. The co-rotating flow (yellow) and southward magnetic field (blue) induce an electric field directed away from Jupiter (green), which drives currents (purple) across Io. 6

1.3 Three models of the Io interaction. Top, the steady current loop of Goldreich and Lynden-Bell (1969). Middle, the Alfvén wing model, with the Alfvén wave reflecting off Jupiter's ionosphere and returning to Io's wake well after the field lines have lost contact with Io. Bottom, the interaction proposed by Crary and Bagenal (1997), with the initial Alfvén wave returning while the field lines are still connected to Io, and the currents settling into a steady state after several Alfvén travel times. 9

1.4 The discovery image of Io's auroral spot. The image was taken with the Infrared Telescope Facility's PROCAM instrument. Top: The raw image; Middle: A smooth fit to the main auroral oval; Bottom: The image with the smooth fit subtracted. [Connerney et al., 1993] 12

1.5 A 3.4 μm image of the jovian aurora taken with the Infrared Telescope Facility's NSFCAM instrument. The Io spot is easily seen at both the north and south foot points [Connerney et al., 1998] 14

1.6 A Hubble Space Telescope image of the jovian aurora. This image was taken in the ultraviolet using the FOC instrument. In the left image, an auroral spot can be seen at the southern foot of the Io flux tube. In the right image, the foot of the Io flux tube is behind the limb, but emission would have been visible if the auroral spot lagged more than 7° behind the instantaneous position of the flux tube [Prangé et al., 1996]. The tilt of Jupiter's magnetic pole places the northern (southern) aurora behind the limb in the left (right) image. . . . 14

1.7 A Hubble Space Telescope image of the jovian aurora. This image was taken in the ultraviolet using the WFPC2 instrument. Emission at the foot of the Io flux tube can be seen equator-ward of the southern oval. In this image, the south pole is tilted towards the Earth. (Image provided by J.T. Clarke.) . . . 15

1.8 A Hubble Space Telescope image of the jovian aurora. These images were taken in the ultraviolet with the new STIS instrument. The two colors show H and H_2 emission. The short lifetime of these lines requires an extended region of particle precipitation, extending approximately 90° along Io's L-shell. 16

1.9 An image of the jovian aurora taken by the Galileo spacecraft's SSI instrument (clear filter). Local midnight is close to the limb in this image. The foot of the Io flux tube is not on the imaged side of Jupiter. 16

1.10 The wave forms of the magnetic fluctuations observed near Io by the Galileo spacecraft, from Huddleston et al. (1997). 18

1.11 The density of the five major species along Io's flux tube. At low latitudes, S^+ (yellow), S^{++} (orange), O^{++} (dark blue) and especially O^+ (light blue) dominate the plasma. Farther along the field line, in the mid- and high latitudes, protons (red) are the only significant species. 23

1.12 The average mass to charge ratio of the plasma, as a function of distance along the Io flux tube. 24

1.13 The Debye length as a function of distance along the Io flux tube. 25

1.14 The electron plasma and electron cyclotron frequencies, as a function of distance along the Io flux tube. Note that the plasma frequency is greater in the torus, but lower in the mid- and high latitudes. 26

2.1 The value of $\omega/\frac{\partial V_A}{\partial z}$ as a function of latitude, for three frequencies. The WKB approximation is valid if this quantity is much greater than 1. 39

2.2 Alfvén velocity according to the diffusive equilibrium models (solid line) and according to an exponential approximation (dashed line). For most of the mid-latitude region, the velocity is approximately exponential with distance along the field line. The speed of light is shown as a dotted line. 41

2.3 Alfvén scale height z_0 calculated from the diffusive equilibrium model. Between 1.5 and 6 R_J, the scale height is nearly constant, and the Alfvén velocity is well-approximated by an exponential. 43

2.4 Rate of change in Alfvén scale height. If $\partial z_0/\partial z \ll 1$, the results of wave propagation in an exponential medium may be applied with reasonable accuracy. 44

2.5 Ratio of the maximum relative velocity for reflected electrons, V_e to that of the Alfvén wave, V_A. Since V_e/V_A is always a small number, the reflected electrons will frequently collied with the Alfvén wave. 52

2.6 The parallel electric potential generated by the Io-genic Alfvén wave because of finite electron inertia. 54

2.7 Rate of particle trapping by Fermi acceleration, per meter of distance along the Io flux tube, as a function of distance from Io. 55

2.8 Energy spectrum of accelerated electrons 56

2.9 Rate of accelerated electron decoupling from the Alfvén wave, per meter of distance along the Io flux tube, as a function of distance from Io. 57

3.1 The components of the electron distribution in a double layer. The electrons entering from the low potential side, f_-, are shown in blue and are accelerated by the electric fields. Those entering from the high potential side, f_+, are shown in green and are slowed. Depending on their initial energy, they may either be reflected or stream through the double layer. 73

3.2 An example of a double layer solution. The double layer is constructed with a potential of $e\Phi_0 = 2kT_e$ across it. The Sagdeev potential, V, the real potential, Φ, the charge density and the ion and electron densities are shown as a function of position. 77

3.3 A series of double layers, showing the transition from a single double layer to multiple layer solutions. The Sagdeev potential and charge density are shown for eight different values of Φ_0. $e\Phi_0 = \{2, 5, 6, 7\}\, kT_e$. 78

3.4 A continuation of figure 3.3 showing double layers with $e\Phi_0 = \{8, 9, 10, 10.95\}$ kT_e. Note the change in the horizontal scale from figure 3.3 79

3.5 The critical current for driving the Buneman instability and causing double layers, as a function of distance along the Io flux tube. The solid line is for $T_e + T_i = 65$ eV, the dashed line for 10 eV. 84

3.6 The velocity profile of the wake. Assuming double layers with a maximum current of $10^{-7} A/m^2$, solid line, and no double layers, dashed line. 88

4.1 Properties of the ion cyclotron waves observed by Galileo in Io's exosphere, from Warnecke et al. [1997]. From top to bottom: Q, the ratio of the intermediate to minimum eigenvalues of the covariance matrix, a measure of the how "wave-like" the observed magnetic field was; ϑ the angle between the background field and the minimum variance direction (i.e. the direction of $\vec{k}$; P, the degree of polarization, E the ellipticity; F, the peak frequency, along with the S^+,SO^+ and SO_2^+ gyrofrequencies; A, the amplitude; and u, the field-aligned component of velocity for pick-up ions (solid line) and the Galileo spacecraft(dotted line). 93

4.2 Spectra of ion cyclotron waves, from Warnecke et al. [1997]. 95

4.3 Contours of a ring-beam (pick-up) distribution, for $a_{\|}$=$a_{\perp}$ = 2.5 km/s, v_b = 7.5 km/s, v_r = 56.5 km/s, consistent with a 5 eV plasma and field line tilt at the time of the Galileo encounter. Contour levels are of 0.1, 0.01 and 0.001 times the peak phase space density. 96

4.4 A cartoon of particle scattering by two waves. Solid lines show the path a particle might follow through phase space, while dotted lines show wave characteristics, along which the particles are scattered. The parallel propagating wave is shown in red while the anti-parallel in blue. Left, from Terasawa [1989], is the situation when the phase velocity is small compared to the particle velocity; right, the case for a large phase velocity. 102

4.5 The spectrum Galileo observed between 17:35-17:40 SCET, solid line, and the model spectrum used in test particle integrations, dashed lines. 105
4.6 Top: wave forms of the ion cyclotron waves observed by Galileo, from Huddleston, 1997. Bottom: Model wave forms used in the test particle integrations. 106
4.7 The dispersion relation for the ion cyclotron waves used in the test particle integrations. 108
4.8 The phase velocity for the ion cyclotron waves used in the test particle integrations. 109
4.9 The group velocity for the ion cyclotron waves used in the test particle integrations. Dashed lines indicate a group velocity which is anti-parallel to the phase velocity. 110
4.10 A phase space plot showing the results of a test particle scattering code. The dashed line shows the characteristic direction for the waves used in the simulation. 112
4.11 The rms parallel velocity of the test particles as a function of time. The dashes line is the velocity predicted by quasi-linear diffusion. with a 0.0001 a^2 offset. 113
4.12 The rms parallel velocity of the test particles after 50 Ω_c^{-1} seconds, for several wave amplitudes. The line is the velocity predicted by quasi-linear diffusion. 114
4.13 Velocity space distribution of the S^+ test particles at t = 0 (top) and t = $10\Omega_c^{-1}$. The particles which were originally in the Maxwellian core are shown in blue and those originally in the ring-beam distribution are shown in red. . 115
4.14 As figure 4.13, but for t = $25\Omega_c^{-1}$ (top) and t = $50\Omega_c^{-1}$ (bottom). 116
4.15 Parallel temperature of the test particles, for particles which were originally part of the Maxwellian population (solid line) and part of the ring-beam population (dashed line). 117
4.16 Thermal anisotropy of the test particles, for particles which were originally part of the Maxwellian population (solid line) and part of the ring-beam population (dashed line). 118
4.17 The upper limit on O^+ pick-up ion density, based on marginal stability of ion cyclotron waves. The background, O^+ density is 2000 cm^{-3}. 120

4.18 The geometry of the Galileo JO encounter with Io. The solid line shows the spacecraft's trajectory, with the thick line showing the regions in which ion cyclotron waves were observed. The dashed line shows the path integral used in equation 4.30, an approximate stream line. The dashed and dotted lines together show the line of sight for a terrestrial observer. 123

4.19 The SO_2^+ density profile of Huddleston et al., 1997, black. Colored lines show the predictions of various models. Extrapolating from the neutral Na density [Schneider, 1988] gives the purple ($Na:SO_2$ = 1:100) and green ($Na:SO_2$ = 50). The outflow model of Scherb and Smyth, 1993, results in the yellow (1.4 and 1 hour dissociation times) and orange (0.5 hour dissociation times). . . 127

Tables

Table

1.1 Assumed plasma conditions at the time of the Galileo J0 encounter 27

4.1 Charge exchange rates for Io's exosphere, from McGrath and Johnson, 1989. . 124

Chapter 1

Introduction

1.1 Overview

Jupiter's inner most major moon, Io, interacts strongly with the jovian magnetosphere, and generates energetic, high latitude phenomena near Jupiter. This interaction was first discovered in 1964, when Bigg noted that jovian decametric radio emissions occur primarily when Io is at certain orbital phases. Figure 1.1 is a sketch showing the elements of the Io interaction, as we currently understand them. In the high latitudes, the interaction generates strong radio emissions between a few and 40 MHz and an auroral feature, probably caused by the precipitation of energetic electrons. Since their discovery, the Io-controlled decametric emissions have been extensively studied [Genova et al., 1989, and references therein] and many theoretical models of the emission process have been suggested [Goldstein and Goertz, 1983, and references therein]. In the past five years, observations of the jovian aurora have revealed emission near the foot of Io's flux tube, in the infrared [Connerney et al., 1993], ultraviolet [Clarke et al., 1996] and visible (Andrew Ingersoll, personal communication, 1998). The visible emissions can only be detected on the night side of Jupiter, and therefore have only been seen by the Galileo spacecraft.

Near Jupiter's equator, Io orbits through a dense region of plasma. The plasma has a density of approximately 2500 cm^{-3} and is primarily composed of sulfur and oxygen ions. The source of these ions is volcanic activity on Io. The plasma co-rotates with Jupiter, with a velocity of 74 km/s at Io's distance from Jupiter. Since Io's orbital velocity is 17 km/s, there is an induced, vxB, potential across Io of 0.113 V/m. Io is electrically conductive, and this induced potential drives strong currents. The Voyager 1 spacecraft observed a $2x10^6$ Amp current in the Io flux tube [Acuña et al., 1981]. These currents may close through field-aligned, steady (i.e. Birkeland) currents and the ionosphere of Jupiter, or they may close in the plasma itself at the front of an Alfvén wave [Hill et al., 1983, and references therein]. The former corresponds to a steady current, and the later a transient disturbance. It is not clear which of these descriptions is more accurate, although Galileo data suggests that the interaction involves some combination of steady and transient currents [Crary and Bagenal, 1997].

Relatively little work has been done on coupling the processes at the two ends of the Io flux tube. The interaction near Io is modeled as an MHD process (e.g. Neubauer

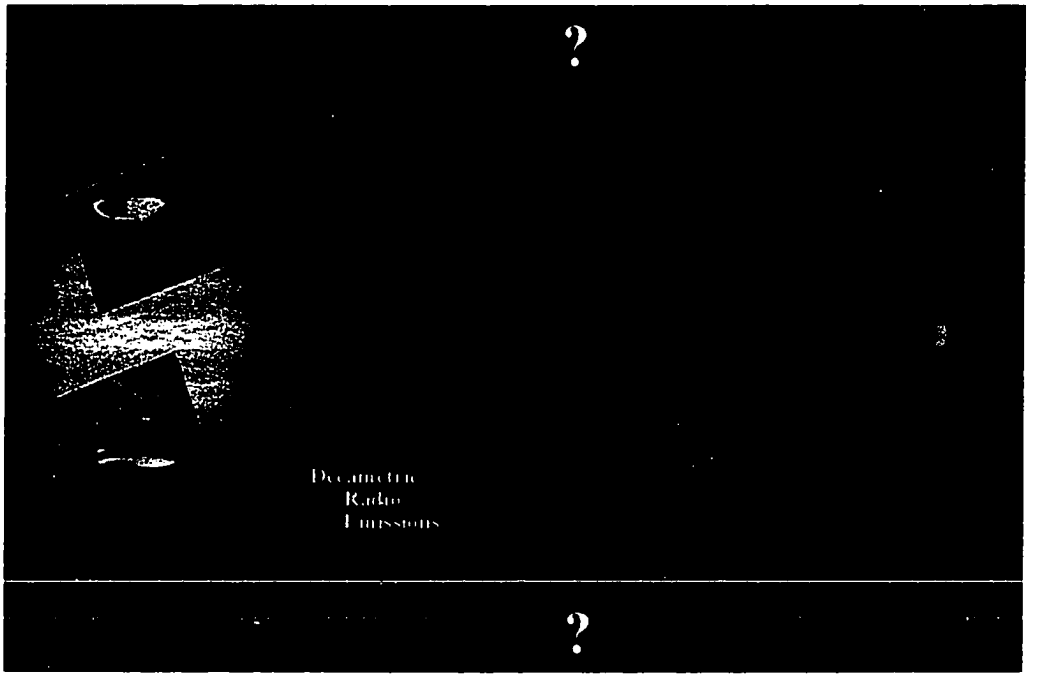


Figure 1.1: Overview of the Io's interaction with the jovian magnetosphere. Plasma density is shown in purple, and based on the Bagenal (1994) model. HST/STIS images of the jovian aurora and Io auroral spot and Galileo SSI image of Io.

1980; Linker et al., 1997) while theories of the radio emissions are kinetic. These theories and analysis of auroral observations typically assume the presence of accelerated electrons (e.g. Goldstein et al., 1983, Zarka et al., 1996, Prangé et al. 1996).

The nature of the Io interaction and how it results in high latitude phenomena is very similar to many problems in terrestrial, auroral physics. The fundamental issue is how a source of steady Birkeland currents and/or Alfvén waves generate energetic particle precipitation and radio emissions near the electron cyclotron frequency. In the case of Alfvén waves, there is the additional problem of wave propagation: The plasma properties, specifically the Alfvén velocity, vary by orders of magnitude between the source and the high latitudes, implying significant reflection of wave energy. In the case of the jovian magnetosphere, the source of the disturbance is Io; in the case of the terrestrial magnetosphere, the source is transient and located somewhere in the magnetotail (e.g. reconnection along a near-Earth neutral line.) In both cases, however, the physical processes are similar.

Despite much more limited data, the Io interaction is in some ways easier to study that the terrestrial equivalent. The location of the source is well known, allowing spacecraft to unambiguously measure the source and its environment. The Voyager 1 spacecraft passed roughly one Io radius upstream of the affected flux tube, and measured the magnetic field and flow velocity perturbations from the disturbance in the far field [Acuña et al., 1981; Barnett and Olbert, 1986]. The Galileo spacecraft, on its initial (J0) orbit around Jupiter, passed 900 km directly downstream of Io [Kivelson et al., 1996; Frank et al., 1996; Williams et al., 1996; Gurnett et al., 1996] although the implications of these observations are still a matter of debate. On its extended mission, Galileo will pass 500 km directly up stream of Io and 300 km over Io's south pole, passing directly through the affected flux tube. The Voyager and Galileo spacecraft have also measured the composition, temperature, flow velocity, etc. of the plasma near Io's L-shell [Bagenal and Sullivan, 1981; Bagenal, 1994; Frank et al., 1996; Crary et al., 1998] and additional measurements will be made during Galileo's extended mission. The plasma environments near the source and the high latitude phenomena are constant with time and longitude, to within a factor of a few. This allows reasonable comparisons between observations of the source and high latitude phenomena, even when these observations are not simultaneous.

Io's plasma environment is also a result of Io's magnetospheric interaction. The Io

plasma torus is primarily composed of sulfur and oxygen ions, and the ultimate source of this plasma is Io's neutral atmosphere. The details of the escape processes, and therefore the nature of the Io torus, are poorly understood but probably related to the details of Io's magnetospheric interaction [Wilson, 1996, and references therein]. The Galileo observations of strong ion cyclotron waves near Io provides some insight into the loss process and the source of the Io torus.

In this thesis, my goal is to examine the particle acceleration and scattering processes which result from Io's interaction with the jovian magnetosphere. I present models of the coupling between Io and Jupiter which are consistent with observations and which explain how the MHD interaction near Io produces energetic electron beams in the high latitudes. I then consider the role of ion scattering by the ion cyclotron waves observed near Io. I use these data to estimate (or at least place limits on) the magnitude of the O^+ and S^+ source in Io's exosphere, the temperature and anisotropy of the newly-ionized population, and the neutral column density of Io's exosphere.

In this introduction, I describe the main theories of the interaction near Io, the relevant properties of Jupiter's decametric radio emissions and the Io-related auroral features, Galileo measurements of strong ion cyclotron waves near Io, and the models and observations of plasma conditions near Io which I use in my analysis.

1.2 The character of Io's electrodynamic interaction

Io's orbital motion generates an induced electric potential and strong, field-aligned currents. This interaction between Io and Jupiter and its magnetosphere has been described as either a steady, uniform current loop, closing through Io and Jupiter's ionosphere [Goldreich and Lynden-Bell, 1969], an Alfvén wave generated by Io with no coupling to Jupiter's ionosphere and current closure in the ambient plasma [Goertz, 1980; Neubauer, 1980] or as a combination of the two [Crary and Bagenal, 1997]. Figure 1.2 shows the geometry of the of the interaction near Io. The models have different implications for Io's generation of high-latitude phenomena. The nature of the interaction depends on the time scale required to establish an uniform current loop: The Alfvén wave model is the limiting case where flux tubes are advected past Io much more rapidly than the system can settle into a steady state;

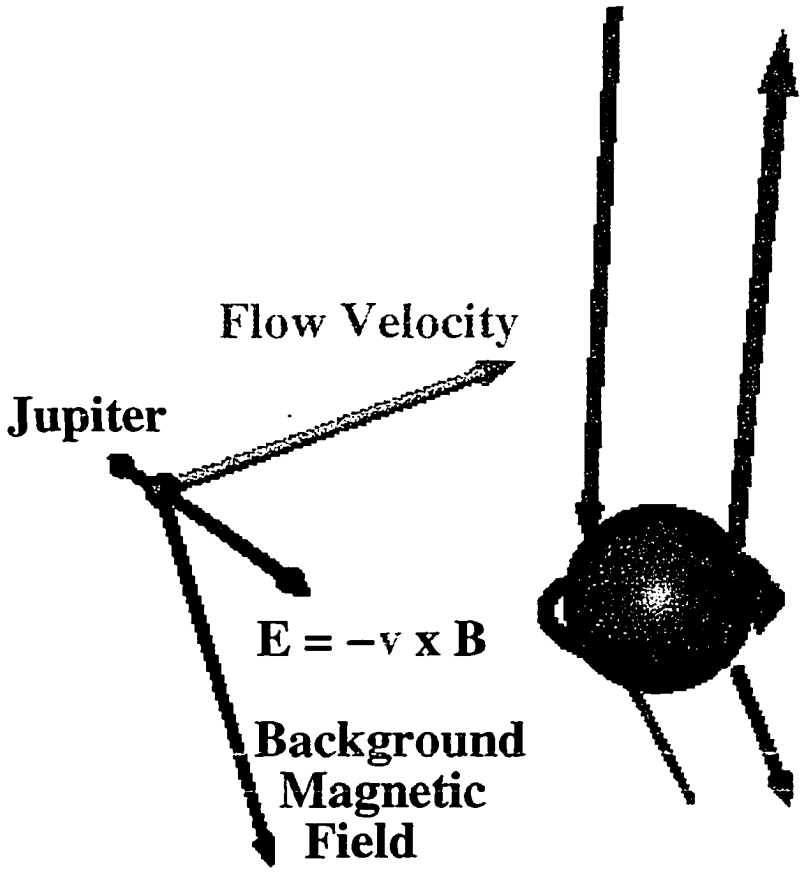


Figure 1.2: A sketch of the Io interaction's geometry. The co-rotating flow (yellow) and southward magnetic field (blue) induce an electric field directed away from Jupiter (green), which drives currents (purple) across Io.

the uniform current loop is the limiting case where a steady state is reached rapidly. Here, we use "steady state" to mean that $\frac{\partial}{\partial t} = 0$ along a given field line. The time scale to reach a steady state is determined by Alfvén wave propagation [Goldreich and Lynden-Bell, 1969; Neubauer, 1980].

For currents to close in Jupiter's ionosphere, the disturbance generated by Io must reach Jupiter; for currents driven in Jupiter's ionosphere to affect the interaction near Io, they must propagate back to Io. This disturbance propagates at the Alfvén velocity. Based on the Voyager 1 observations of the plasma near Io, the round-trip, Alfvén propagation time is 950 sec, when Io is in the center of the plasma torus. Using the higher electron density observed by Galileo, 3770 cm^{-3} [Gurnett et al., 1996; Frank et al., 1996] but assuming Voyager epoch composition and temperatures, gives a propagation time of 1212 seconds (450/1970 seconds) when Io is in the center (above/below) the torus. The unperturbed advection (i.e. the diameter of Io divided by the upstream flow velocity) is approximately 60 seconds, much less than the Alfvén propagation time. However, Galileo also observed a nearly-stagnated flow in Io's wake, with a velocity under 1 km/s [Frank et al., 1996]. This implies that the advection time could be as long as 3600 seconds. Thus, for the first 1200 seconds after the flux tube encounters Io, the interaction between Io and flux tube is that an Alfvénic disturbance which is stationary in Io's rest frame, as in the analytic models of Goertz [1980] and Neubauer [1980].

A single propagation time is insufficient to reach the steady state of Goldreich and Lynden-Bell. The interaction continues to evolve as the Alfvénic disturbance is partially reflected first by Jupiter and then Io [Goertz and Deift, 1973]. Since the disturbance is only partially reflected, each subsequent reflection is of lower amplitude, and results in a lesser change to the current system. The amplitude of the disturbance after n round-trip travel times may be expressed as a convergent series

$$\Delta B_n = \Delta B_0 \left[1 + \left(r_{Io}^{-1} + 1 \right) \sum_{m=1}^{n} \left(r_J r_{Io} \right)^m \right] \tag{1.1}$$

where ΔB_n is the amplitude and r_J and r_{Io} are the reflection coefficients for Alfvén waves, at Jupiter's ionosphere and at Io, calculated from the Alfvén conductivity and the conductivity of Jupiter or Io using eq. 10.10 of Hill et al., 1983. A number of reflections are required to

achieve a Goldreich and Lynden-Bell steady state. The number depends on the conductance of Io and Jupiter and the properties of the plasma. For estimated values in the range $0.2 < \Sigma_J < 4\mho$ [Strobel and Atreya, 1983] and $30 < \Sigma_{Io} < 200\mho$ [Summers and Strobel, 1996], I find that after 2 to 5 propagation times, or 1200 to 4800 sec. after the return of the initial wave, the interaction is within 5% of the steady state.

A flux tube encountering Io would carry an initial, Alfvénic current, of the sort modeled by Goertz and by Neubauer, for roughly the first 1200 sec. after contacting Io. For another 1200 to 4800 sec., the flux tube would be in transition, and contain an Alfvén wave reflecting back and forth between Io and Jupiter while gradually decreasing in amplitude. After this transition period, if the field lines remain in contact with Io, the flux tube would reach the steady state. The Galileo spacecraft's PLS instrument measured flows of under 1 km/s in Io's wake [Frank et al., 1996]. This means some of the field lines would spend more than 3630 seconds interacting with Io, and will reach a steady state. The field lines would remain in this steady state until advecting past Io. Then, they would be accelerated back to corotation, at a rate determined by plasma inertia and Jupiter's ionospheric conductance, producing an extended wake. Figure 1.3 shows cartoons of the steady current, purely Alfvénic, and transitional models of the interaction.

1.3 Io-related radio emissions

Io's importance to the jovian magnetosphere was discovered in 1964, when Bigg noted that Jupiter's decametric radio emissions (DAM) were strongly controlled by the orbital phase of Io [Bigg, 1964]. These emissions are at frequencies between a few and 39.5 MHz. The lower limit is poorly determined, in part because it is only observable by spacecraft and in part because other, morphologically different jovian emissions (e.g. hectometric emissions or HOM) are also present at a few MHz. The upper limit corresponds to the peak electron cyclotron frequency along Io's L-shell.

The decametric emissions are intense, with a peak flux of up to $3\text{x}10^7$ Jy at near 10 MHz, in comparison to the few Jy peak flux around 1 GHz for decimetric (synchrotron) emission from Jupiter's inner radiation belts. (1 Jy = 10^{-26} W/m^2/Hz.) If this were emitted isotropically, the total power output in the decametric band would be approximately $3\text{x}10^{11}$

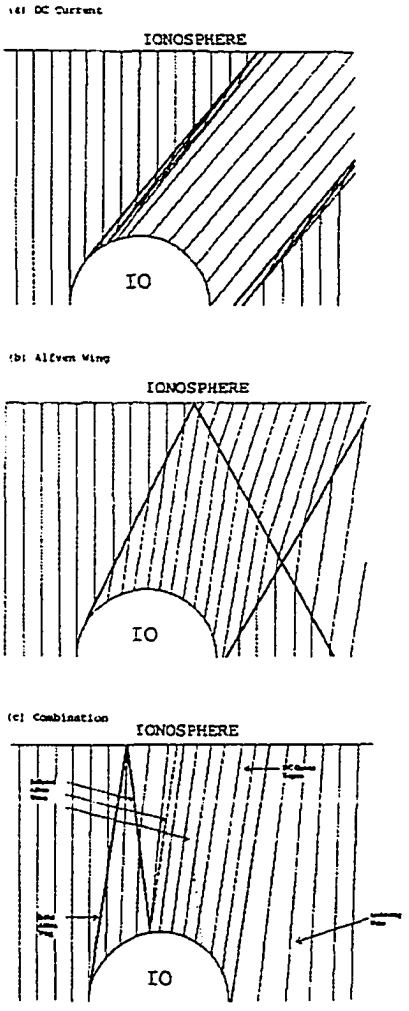


Figure 1.3: Three models of the Io interaction. Top, the steady current loop of Goldreich and Lynden-Bell (1969). Middle, the Alfvén wing model, with the Alfvén wave reflecting off Jupiter's ionosphere and returning to Io's wake well after the field lines have lost contact with Io. Bottom, the interaction proposed by Crary and Bagenal (1997), with the initial Alfvén wave returning while the field lines are still connected to Io, and the currents settling into a steady state after several Alfvén travel times.

W [Carr et al., 1983]. However, isotropic emission is unlikely. The Io control of the emissions and statistical analysis of observation probabilities suggests that the emission is restricted to a thin, hollow cone centered on the magnetic field line. The data suggest the beaming cone has a half-angle of approximately 70-80° and a thickness of a few degrees (originally noted by Dulk, 1965, more recent work is reviewed by Zarka, 1988) and emission at of slightly above the local electron cyclotron frequency, f_{ce}. The beaming implies an emitted power of approximately 5×10^9 W. Decametric "storms" typically last for a few hours when the magnetic field geometry and the location of Io places the Earth within the beaming cone. Most of these emissions, "L-bursts", are modulated on periods a few minutes and longer. In addition, "S-bursts" are observed during decametric storms. S-bursts are characterized by short duration bursts, with a bandwidth of a few hundred kHz, which last for a few to a few tens of milliseconds and have rapidly decreasing frequencies. These are interpreted as a beam of electrons, which have recently mirrored off the atmosphere of Jupiter and are moving upward in a loss cone distribution. The rapidly decreasing frequency is a result of the electrons moving into a region of lower magnetic field strength, and therefore lower f_{ce} as they travel away from Jupiter [Carr et al., 1997; Zarka et al., 1996]. S-burst emission represent roughly 10% of the emitted, decametric power [Carr et al., 1983; Kaiser, 1989].

The waves' polarization [Kaiser, 1989], strong amplitude and beaming of the decametric emissions suggest that they are X-mode waves generated by the cyclotron maser instability [Genova et al., 1989; Zarka et al., 1996] and analogous to the Earth's auroral kilometric radiation [Wu and Lee, 1979; Menietti and Burch, 1993, and references therein] The cyclotron maser instability occurs when electrons have a loss cone distribution. R-X mode waves at the local upper hybrid frequency ($f_{UH} = \sqrt{f_{ce}^2 + f_{pe}^2} \sim f_{ce}$ since $f_{ce} \gg f_{pe}$) cause electrons to scatter into the loss cone. In the process, the waves are coherently amplified. This emission process has a saturation efficiency of a few percent [Lee, 1989], i.e. a few percent of the electrons' energy is converted to radio emissions. As a result, this mechanism requires a $\sim 5 \times 10^{11}$ W source of energetic electrons in a loss cone distribution. By analogy to the terrestrial kilometric emission, the source electrons are generally assumed to have energies of a few to a few tens of keV.

The location of the Io-related source of radio emissions is not well known. The frequencies correspond to the electron cyclotron frequencies at the high-latitudes of Io's L-

shell. However, the longitude of the source relative to Io is poorly quantified. Emissions are observed over 50-90° of Io phase. Uncertainties in the emission cone angle, especially the possibility that it may depend on electron energy, pitch angle and frequency, make an accurate determination of the source difficult. Estimates vary, but in general suggest that the source of radio emissions extends downstream of Io for tens of degrees. For example, Leblanc et al. [1994] assumed a relation between the degree of elliptical polarization and the emission cone angle, and inferred decametric source locations ranging from 31° upstream of Io, for the left hand polarized C source, to 71° down stream of Io, for the left hand polarized B source. For causal reasons, an up stream source is unlikely, and probably a result of uncertainties in the magnetic field model and/or the assumed emission cone angle. However, Leblanc et al. clearly show that the radio source is not localized to Io's instantaneous flux tube.

1.4 Io's auroral spot

The aurora of Jupiter were first observed by the Voyager 1 spacecraft's ultraviolet spectrometer [Broadfoot et al., 1981] and later from Earth and Earth orbit, at ultraviolet, infrared and x-ray wavelengths [Baron et al., 1991; Kim et al., 1991; Waite et al., 1994]. However, none of these observations had sufficient resolution to examine the small scale structure of the aurora, nor to identify any features associated with the Io interaction.

In 1993, Connerney et al. identified an auroral spot located near the foot of Io's flux tube. The discovery was made using the Infrared Telescope Facility and looking at 3.8 μm emissions from H_3^+. This wavelength is particularly convenient for auroral observations, since it is also in a deep methane absorption band. This eliminates most of reflected sunlight from the disk of Jupiter. Figure 1.4 shows the discovery image of Io's auroral spot. Without processing, a slight enhancement can be seen, and the spot is more apparent when a model of the main auroral oval is subtracted. In 31 observations where the spot was clearly detected, Connerney et al. found that it tracked the motion of Io, remaining within approximately 15° of Io's L-shell and led the predicted position of Io's flux tube by 15-20° of longitude. The uncertainty in the magnetic field model used, both for the mapping of the L-shell and instantaneous flux tube is roughly 10-15° [Connerney, 1991].

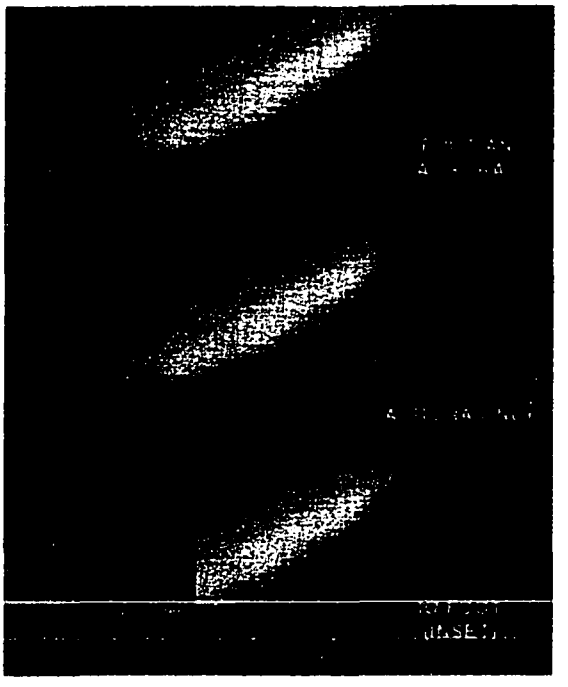


Figure 1.4: The discovery image of Io's auroral spot. The image was taken with the Infrared Telescope Facility's PROCAM instrument. Top: The raw image; Middle: A smooth fit to the main auroral oval; Bottom: The image with the smooth fit subtracted. [Connerney et al., 1993]

Since its discovery, images of Io's auroral spot have improved dramatically. In the infrared, observations of H_3^+ have continued and resolution has improved substantially with the 1994 addition of the NSFCAM instrument, e.g. the image in figure 1.5. Some of these images show evidence of extended emission in an arc along Io's L-shell [Connerney et al., 1995].

In addition to the infrared, Io's auroral spot has been observed by the Hubble Space Telescope using the Faint Object Camera, figure 1.6, Wide Field Planetary Camera 2, figure 1.7 and most recently by the Space Telescope Imaging Spectrograph, figure 1.8. These observations are of ultraviolet emission from H and H_2. In the STIS images, an extensive arc can be seen, beginning at the foot of the Io flux tube and continuing along Io's L-shell in the downstream direction for approximately 90^o of longitude. Since these UV lines have a lifetime of well under a second, particle precipitation must occur along this entire arc, which maps back to Io's wake. In this thesis, I use the terms "Io spot" and "Io related arc" to distinguish between these two features. The main auroral oval is poleward of the Io features, consistent with a source in the middle or outer magnetosphere.

The Galileo spacecraft has also observed the aurora of Jupiter in the visible. Figure 1.9 shows an example, taken through the clear filter (0.4 to 1.0 μm). In addition to higher resolution than the Earth-based images, the Galileo images are of the night side of Jupiter's aurora. In figure 1.9, the foot of the Io flux tube is behind the limb of Jupiter. Other Galileo images, unpublished and not yet released, include the Io spot and resolve brightness structure within it (Andy Ingersoll, personal communication, 1998). Near infrared imaging spectra have been taken with Galileo's NIMS instrument and data analysis is in progress (personal communication, Robert Carlson, 1998). Efforts to obtain UV spectra of the Io spot with Galileo have been unsuccessful (Wayne Pryor, personal communication, 1998) due to the instrument's long slit and observing geometry, which prevent isolating the Io spot from other auroral features.

These observations place several constraints on models of the Io interaction. Based on the size and brightness of the auroral spot, an input power of a few times 10^{11} W is required [Prangé et al., 1996; Clarke et al., 1996] in the form of particles with few to hundreds of keV energy Prangé et al. [1996]. The extent of the arc, in longitude, suggests a source which is physically larger than Io, and probably related to re-acceleration of the wake, which

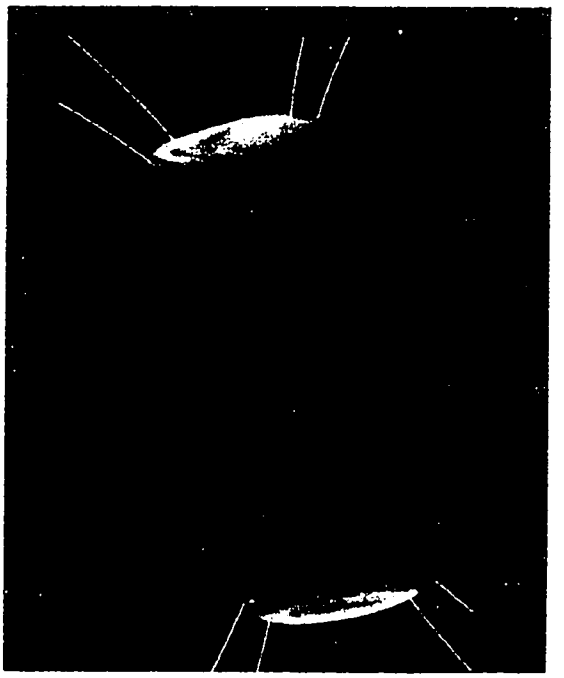


Figure 1.5: A 3.4 μm image of the jovian aurora taken with the Infrared Telescope Facility's NSFCAM instrument. The Io spot is easily seen at both the north and south foot points [Connerney et al., 1998]

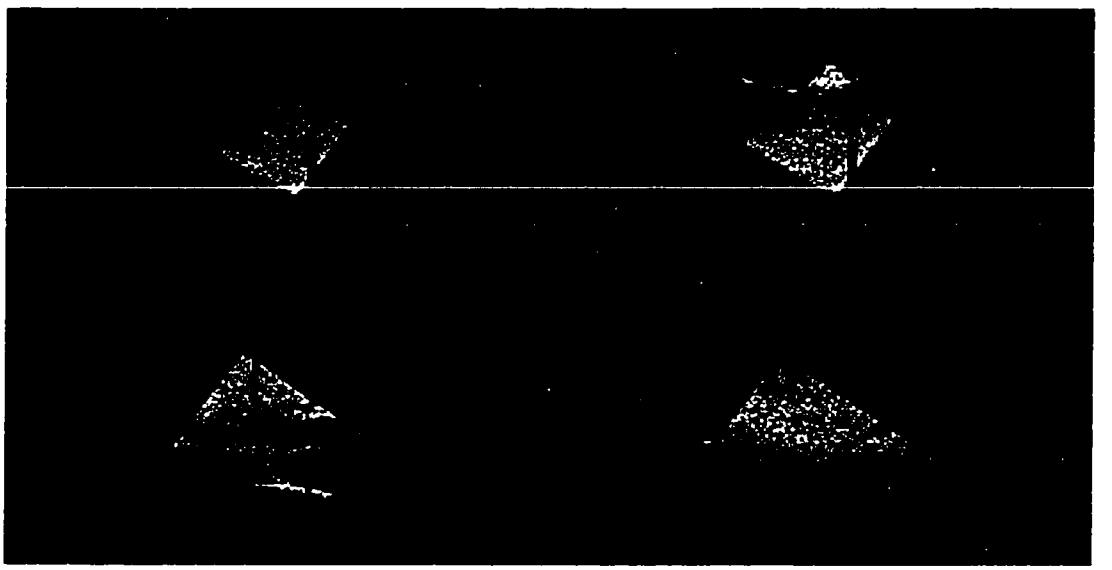


Figure 1.6: A Hubble Space Telescope image of the jovian aurora. This image was taken in the ultraviolet using the FOC instrument. In the left image, an auroral spot can be seen at the southern foot of the Io flux tube. In the right image, the foot of the Io flux tube is behind the limb, but emission would have been visible if the auroral spot lagged more than 7° behind the instantaneous position of the flux tube [Prangé et al., 1996]. The tilt of Jupiter's magnetic pole places the northern (southern) aurora behind the limb in the left (right) image.

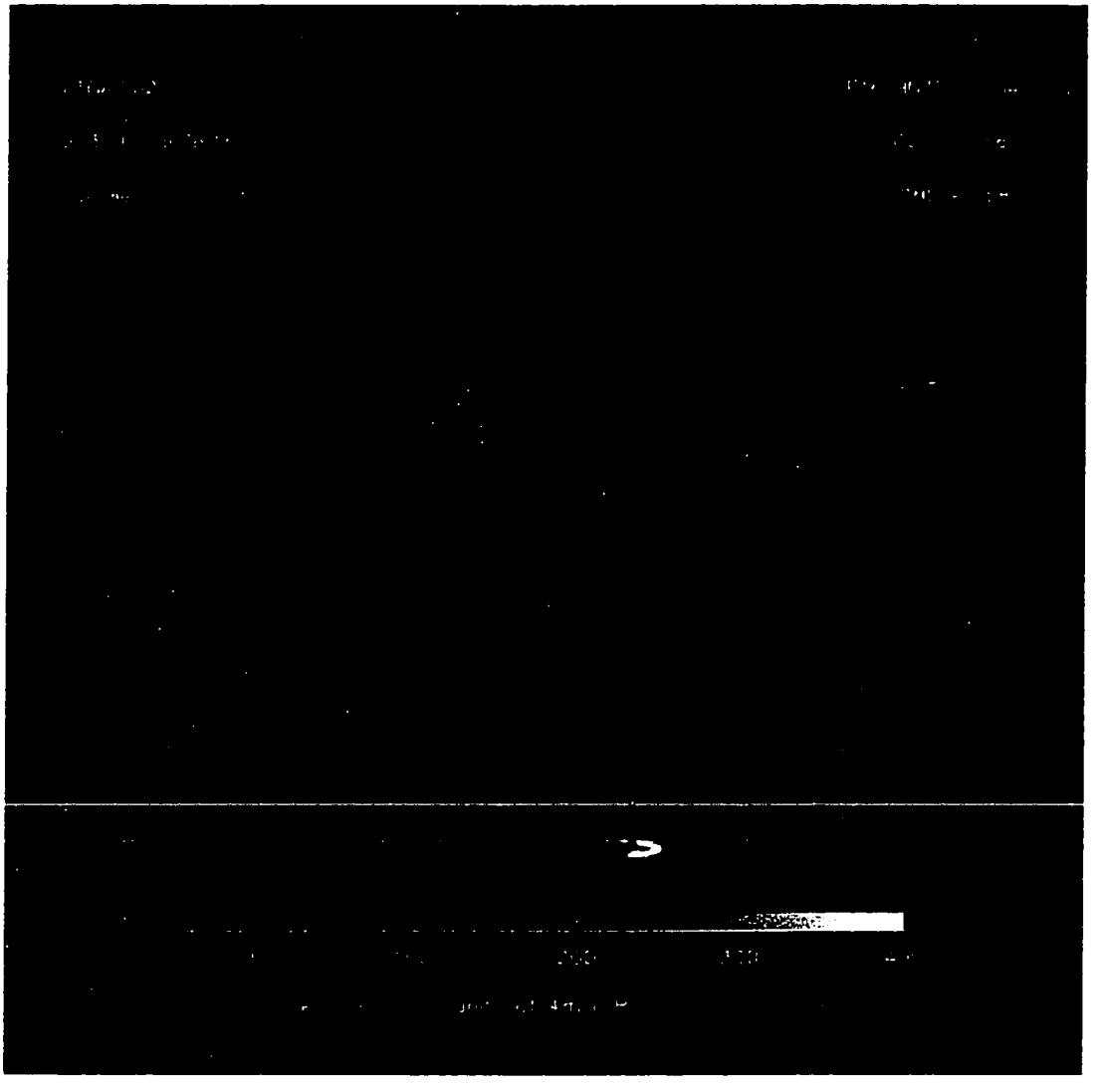


Figure 1.7: A Hubble Space Telescope image of the jovian aurora. This image was taken in the ultraviolet using the WFPC2 instrument. Emission at the foot of the Io flux tube can be seen equator-ward of the southern oval. In this image, the south pole is tilted towards the Earth. (Image provided by J.T. Clarke.)

Figure 1.8: A Hubble Space Telescope image of the jovian aurora. These images were taken in the ultraviolet with the new STIS instrument. The two colors show H and H_2 emission. The short lifetime of these lines requires an extended region of particle precipitation, extending approximately 90° along Io's L-shell.

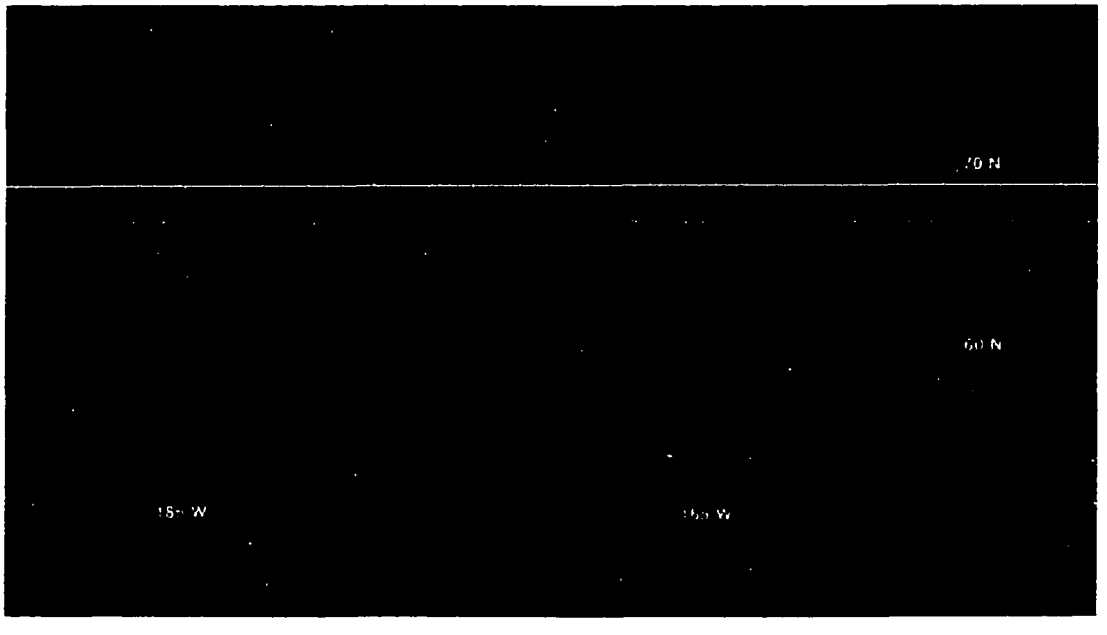


Figure 1.9: An image of the jovian aurora taken by the Galileo spacecraft's SSI instrument (clear filter). Local midnight is close to the limb in this image. The foot of the Io flux tube is not on the imaged side of Jupiter.

was observed by Galileo to be at near-stagnant velocities 900 km downstream of Io. The location and longitudinal offset of the spot itself could provide information on the character of the interaction: The Alfvén wing and steady current models make significantly different predictions. Unfortunately, past magnetic field models are too uncertain. The most recent model, Connerney [1998], is based in part on the observed location of the spot.

1.5 Ion cyclotron waves near Io

On Dec. 7, 1995, the Galileo spacecraft made a close encounter with Io, passing 900 km downstream at closest approach. During the encounter, the magnetometer observed intense magnetic field fluctuations with a period of 2-3 seconds. Figure 1.10, from Huddleston et al., 1997, shows the observed wave forms. These waves had near-circular, left-hand polarization and the perturbations were within 7^o of perpendicular to the background magnetic field. This indicates that they were electromagnetic ion cyclotron waves [Kivelson et al., 1996; Warnecke et al., 1997]. The amplitude of these waves was extremely large, approaching 100 nT, or roughly 6% of the background field strength. The amplitude is asymmetric about Io, falling off more rapidly with radius on the sub-jovian side, and extending to ~ 5 and 12 R_{Io} on the sub- and anti-jovian sides, respectively.

The presence of ion cyclotron waves is not surprising. Newly ionized particles are accelerated by the $-\vec{v} \times \vec{B}$ electric field and form a ring-beam distribution. Both theoretical work and observations of cometary interactions shows that ring-beam distributions are unstable and will generate ion cyclotron waves [Terasawa, 1989, and references therein]. Since Io's exosphere is a known source of ionizations, ion cyclotron waves should be likely. A ring beam distribution can also be subject to the electrostatic, lower hybrid instability. However, this only occurs when the beam component is small, since lower hybrid waves only affect ions with a pitch angle within roughly a degree of 90°. This was not the case during the Galileo/Io encounter due to the $\sim 7^o$ tilt of the magnetic field. The surprising aspect of the observed waves was their frequency. The peak in the spectra is near 0.4 Hz, which is the cyclotron frequency of molecular SO_2^+ ions. No evidence of waves near the S^+ or O^+ cyclotron frequencies was observed. [Warnecke et al., 1997]. SO_2^+ and neutral SO_2 have very short lifetimes in the Io torus, due to dissociative recombination and electron impact dissociation.

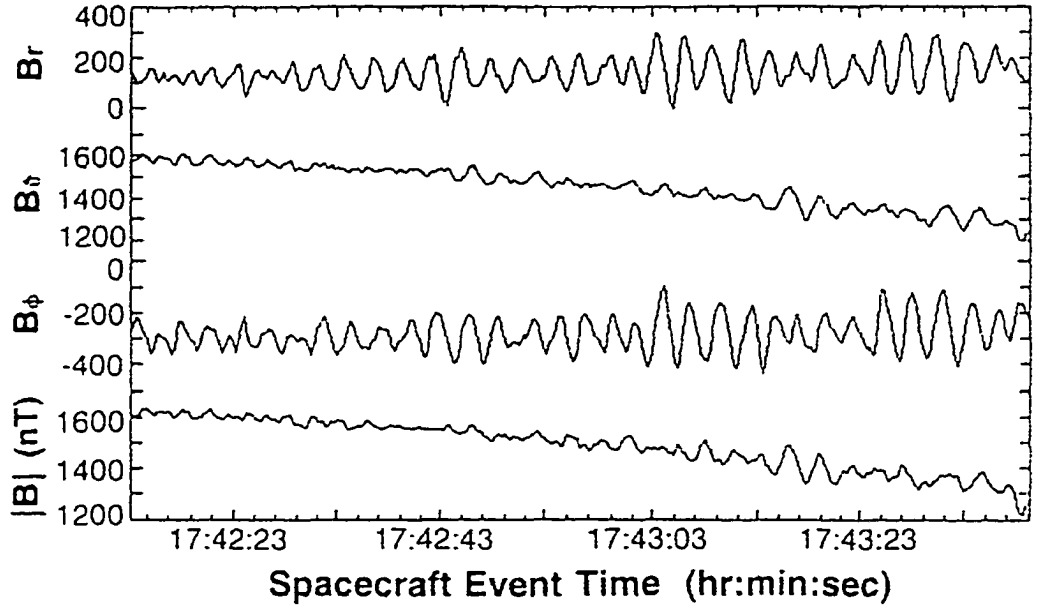


Figure 1.10: The wave forms of the magnetic fluctuations observed near Io by the Galileo spacecraft, from Huddleston et al. (1997).

The exact lifetimes are uncertain, but they are generally believed to be of order one hour, but may vary from 30 minute to 90 minutes, depending on the electron temperature and density.

The presence of SO_2^+ waves, and the absence of S^+ or O^+ waves is, counter-intuitively, due to the absence of thermalized SO_2^+ [Warnecke et al., 1997; Huddleston et al., 1997]. A thermalized background of ions can prevent wave growth, so long as the density of the pick-up ions is sufficiently low and the temperature satisfies certain conditions. The absence of S^+ and O^+ waves could be due to the stabilizing influence of the background plasma, not the lack of a near-Io source. Similarly, SO_2^+ ions dissociate rapidly, and probably never form a thermalized core. Without such a core, any amount of SO_2^+ production would result in ion cyclotron waves. The waves, when combined with the flow velocity of the plasma and certain assumptions, also provide an estimate of the SO_2^+ pick-up ion density at the spacecraft location [Huddleston et al., 1997]. These values range from 1 cm^{-3} far from Io to 100 cm^{-3} at a range of approximately 2 R_{Io}. Closer to Io, strong mirror mode waves were observed, and the ion cyclotron waves either disappeared or were undetectable due to the stronger signal from the mirror mode waves.

1.6 Plasma conditions along Io's flux tube

To model Io-related particle acceleration and scattering, I need to determine (or in some cases, estimate), plasma properties along the Io flux tube. To do this, I used the diffusive equilibrium model [Bagenal, 1994; Huang and Birmingham, 1992; Moncuquet, 1998] and data from the Voyager 1 [Bagenal and Sullivan, 1981; Crary et al., 1996] and Galileo [Gurnett et al., 1996; Crary et al., 1998] spacecraft.

The diffusive equilibrium model allows extrapolations of plasma density, composition and temperature, from a point where these properties are measured to any other point along the same magnetic field line. It is assumed that the plasma along that field line is well mixed. As a result the distribution function of a species is uniform, when expressed in terms of constants of the motion. For example, if a plasma is observed to be multi-species

and Maxwellian at a reference (spacecraft) location, s_0,

$$f(v, s = s_0) = \frac{n_0}{a^3 \pi^{3/2}} e^{-v^2/a^2} \tag{1.2}$$

where n_0 is the density, and a the thermal velocity, the problem is to determine how a and n vary with position. There is generally some potential, U which varies with position (e.g. due to gravitational, centrifugal, or ambipolar electric forces.) Since energy is conserved as the particles move along the field line,

$$U(s) + \frac{1}{2} m_\alpha v^2 = \frac{1}{2} m_\alpha v_0^2 \tag{1.3}$$

where v_0 is the velocity of the particle at the reference location, and where the potential is set equal to zero at this location. It follows that

$$v^2 = v_0^2 - \frac{2}{m_\alpha} U(s) \tag{1.4}$$

$$\begin{aligned} f(v, U(s)) &= \frac{n_0}{a^3 \pi^{3/2}} e^{(v_0^2 - \frac{2}{m_\alpha} U(s)/a^2)} \\ &= f(v, s = s_0) e^{-\frac{2}{m_\alpha} U(s)/a^2} \end{aligned} \tag{1.5}$$

As a result, the temperature of each species is constant and the density of each species is given by

$$n_\alpha = n_{\alpha,0} e^{-\frac{2}{m_\alpha} U(s)/a_\alpha^2} \tag{1.6}$$

Similar relations for density and temperature can be derived for other velocity distributions, e.g. anisotropic Maxwellians [Huang and Birmingham, 1992], κ functions [Meyer-Vernet et al., 1995; Moncuquet, 1998], etc. The only requirements are that the distribution can be expressed in terms of constants of the motion (which must include μ, the magnetic moment, in the case of anisotropic plasmas) and that the resulting distribution function can be integrated to give the density. Unfortunately, functions used as analytic approximations to a ring-beam distribution (e.g. the DGH distribution [Barbosa et al., 1985; Kennel and Engelmann, 1966]) apparently do not result in analytically integrable forms.

Once density is known as a function of potential, the potential must be determined as a function of position. In the case of the Io plasma torus, the dominant forces are the

centrifugal and ambipolar electric potential. Gravity is also significant at high latitudes (close to Jupiter) as is the magnetic mirror force in the case of an anisotropic distribution. The ambipolar field maintains the quasi-neutrality of the plasma. Due to the ion to electron mass ratio, the electron density is far less sensitive to centrifugal and gravitational forces. As a result, an ambipolar field is required to maintain quasi-neutrality. In practice, the centrifugal and gravitational forces are known. For a dipole field,

$$\begin{aligned} U &= U_{C,G} + eZ\Phi_{ambipolar} \\ &= -\frac{GMm}{r} - m\Omega^2 r^2 \sin^2\theta + eZ\Phi_{ambipolar} \end{aligned} \tag{1.7}$$

and a model of the magnetic field may be used to determine position along a given field line. In this thesis, I have used the O6 magnetic field model [Acuña et al., 1983] in combination with the current sheet model of Connerney, Acuña and Ness, 1981. The ambipolar field cannot be calculated analytically, except in the case of a single ion species. To determine densities and compositions, it is necessary to require quasi-neutrality,

$$\sum_{\alpha} Z_\alpha n_\alpha(\Phi_{ambipolar}) = 0 \tag{1.8}$$

and numerically solve for $\Phi_{ambipolar}$.

Using the diffusive equilibrium model, the plasma properties along the Io flux tube may be calculated, given observed or estimated conditions at a given point along the field line. In chapters 2 and 3, I use the conditions observer by the Voyager 1 spacecraft, during its inbound passage through the Io torus [Bagenal, 1994]. (Data from the outbound passage has not been analyzed, except for a brief period around the Io encounter itself, [Barnett, 1986], since the instrument geometry was more favorable during the inbound passage.) Ion density along a given field line peaks at the farthest distance from the rotational axis of Jupiter. The surface defined by these points is called the centrifugal equator. For a tilted dipole field, this is not the same location as the magnetic equator, and is two-thirds of the way between the rotational and magnetic equators. In all cases, I will use plasma conditions for a field line which crosses the equator at 112^o longitude, where the centrifugal and magnetic equators cross. The Voyager observations are extrapolated to this longitude by assuming that plasma properties are constant along a given L-shell as well as a given field line. L-shell is poorly

defined in the Io torus, since magnetic field is not dipolar (i.e. $L = r/\sin^2\theta$ does not define a drift shell.) I define the L of a field line to be the maximum distance, in jovian radii, between the field line and the body center of Jupiter, and calculate this value numerically. This would be identical to the usual definition for a dipole field.

The direct measurements by Voyager did not determine the proton abundance near Io's L-shell, since protons are a minor species in the low-latitudes of the Io torus and because of the instrument's 10 V threshold, nor could the instrument determine the anisotropy of the ions. Some estimate of these are necessary, since the ion anisotropy is critical to determining the vertical extent of the torus, and protons are the dominant species in the mid- and high latitudes. Limits can be placed on the anisotropy and proton density, based on plasma wave observations. During its passage through the torus, Voyager 1 observed a large number of whistlers. The dispersion of whistlers can be used to determine the column density of electrons between the observer and the jovian atmosphere [Menietti and Gurnett, 1980; Tokar, 1982; Crary et al., 1996], and this result may be used to constrain the anisotropy and proton density, although the results are not unique. I use an anisotropy, $T_\perp/T_\parallel$, of 2 and a equatorial proton density of 30 cm^{-3}, roughly in the middle of the range suggested by Crary et al., 1996. This is also consistent with observations of jovian decametric radio emissions [Lecacheux et al., 1991] and more detailed models of whistler propagation [Wang et al., 1998] which indicate low electron densities at high latitudes. The results in chapters 2 and 3 of this thesis are not strongly sensitive to these parameters.

Figure 1.11 shows the resulting ion and electron densities, as a function of distance along the Io flux tube. The heavy ions are confined by the centrifugal potential and remain within one to two jovian radii of the equator. At higher latitudes, protons become the dominant species. Also note that this model does not include ionospheric plasma, so the density at the foot of the flux tube is probably incorrect. The change in composition is also apparent in figure 1.12, showing the average mass to charge ratio of the plasma. In the center of the torus, this is around 17 AMU/e since the most abundant species are S^{++} and O^+, both 16 AMU/e, while at higher latitudes, the ratio drops to the 1 AMU/e value of protons. From the densities and a magnetic field model, the Debye length (figure 1.13) as well as the plasma frequency, f_{pe}, and cyclotron frequency, f_{ce}, (figure 1.14) may be calculated. An interesting property of the plasma and cyclotron frequencies is that they

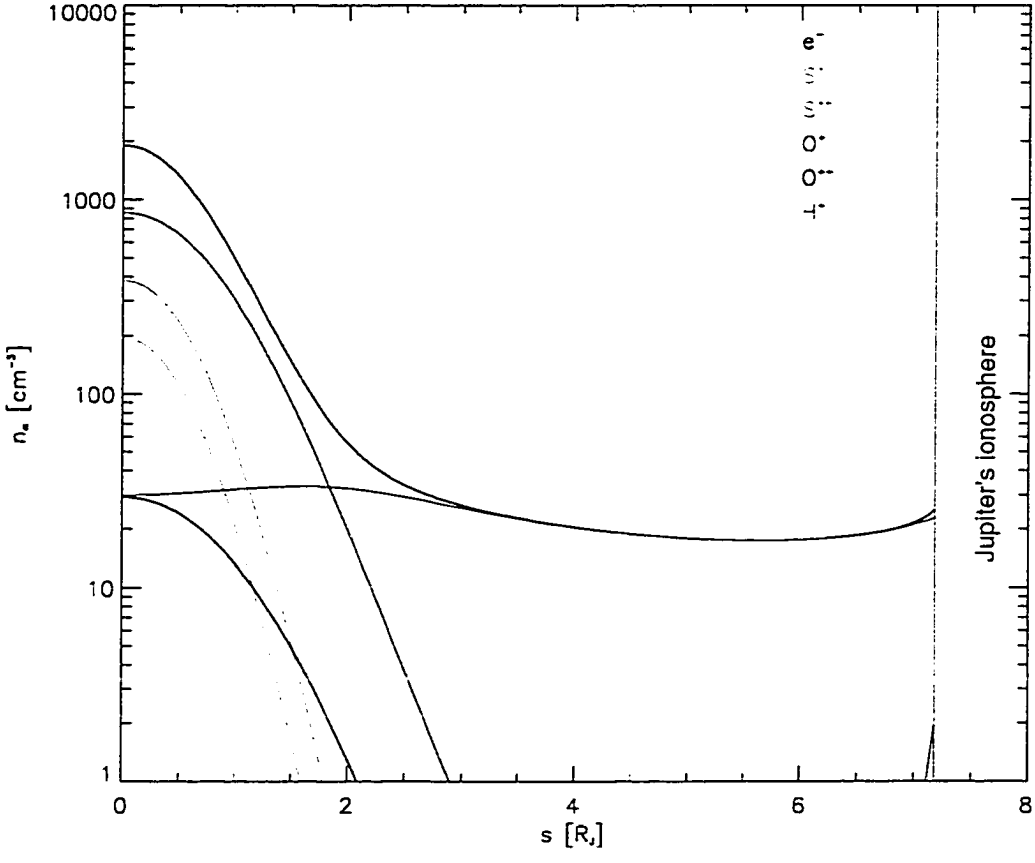


Figure 1.11: The density of the five major species along Io's flux tube. At low latitudes, S^+ (yellow), S^{++} (orange), O^{++} (dark blue) and especially O^+ (light blue) dominate the plasma. Farther along the field line, in the mid- and high latitudes, protons (red) are the only significant species.

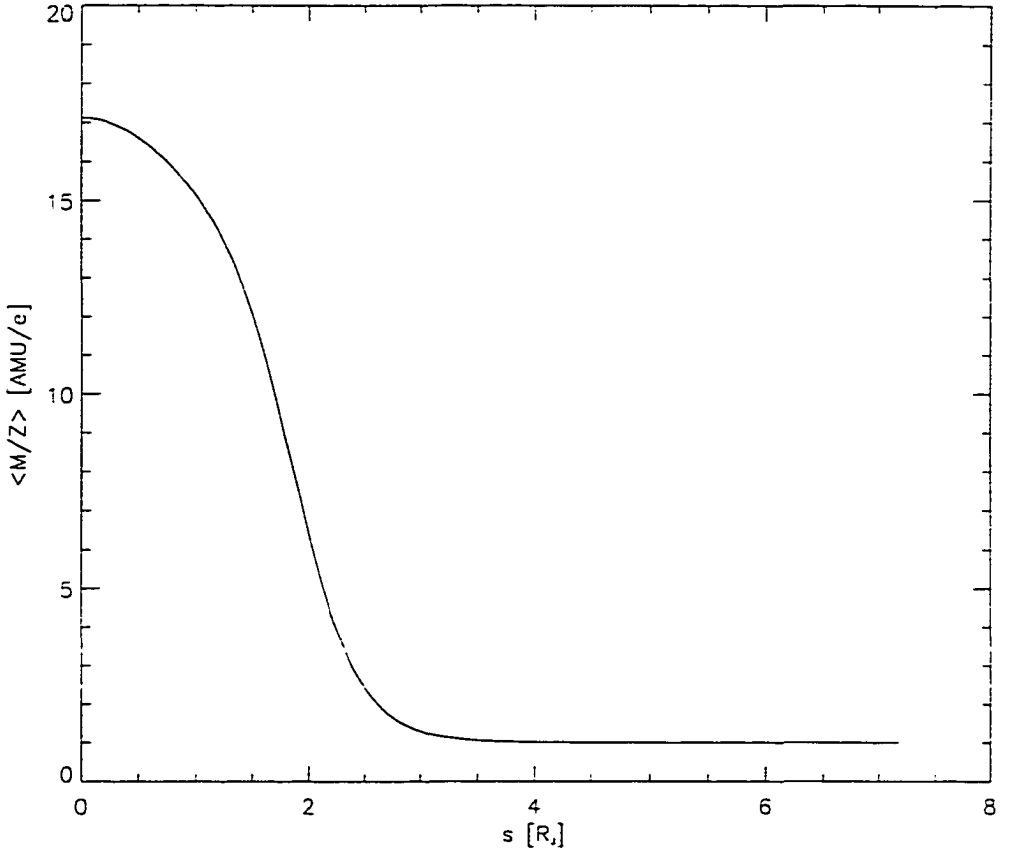


Figure 1.12: The average mass to charge ratio of the plasma, as a function of distance along the Io flux tube.

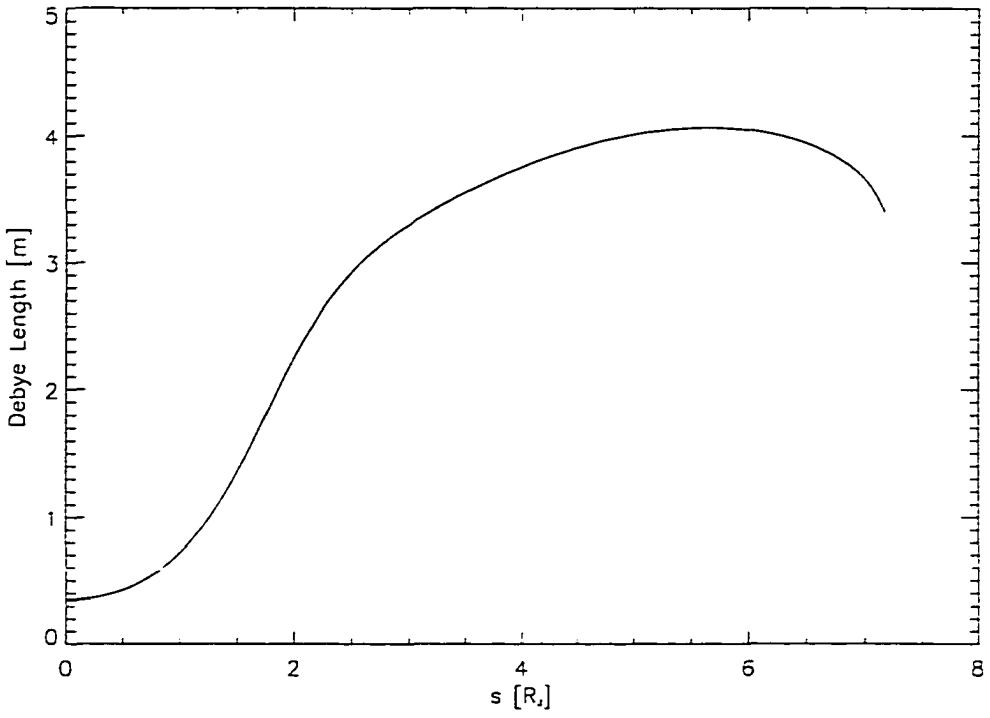


Figure 1.13: The Debye length as a function of distance along the Io flux tube.

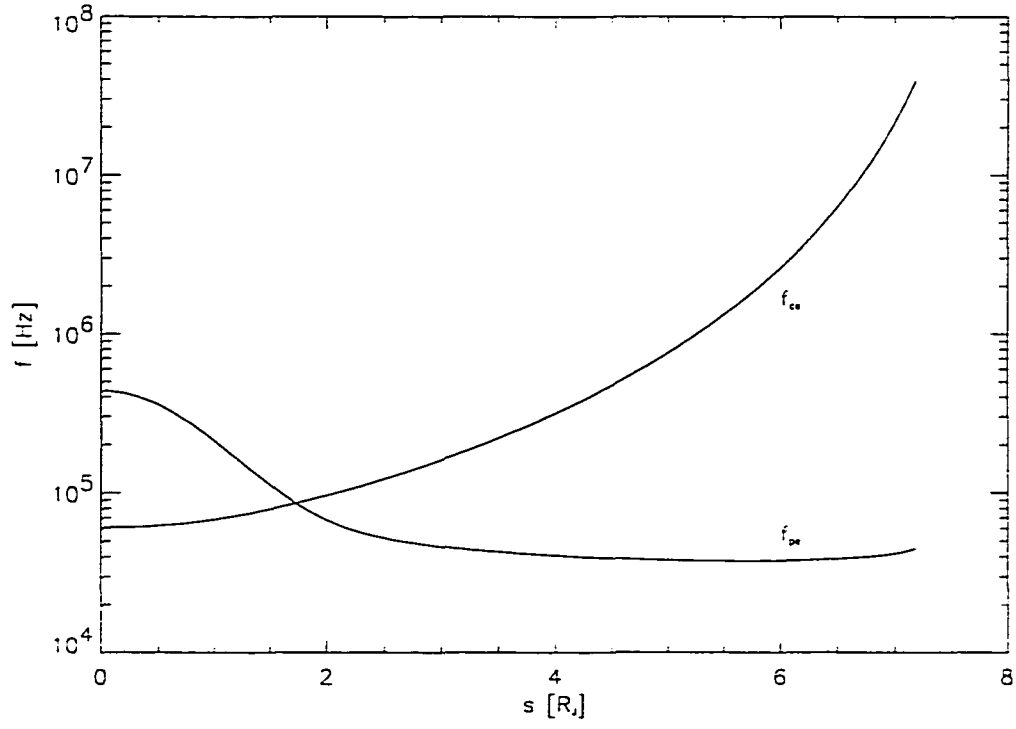


Figure 1.14: The electron plasma and electron cyclotron frequencies, as a function of distance along the Io flux tube. Note that the plasma frequency is greater in the torus, but lower in the mid- and high latitudes.

Species	Density [cm^{-3}]	Temperature [eV]
S^+	260	60*
S^{++}	300	20
O^+	2000	50
O^{++}	460	25
e^-	3780	5†

* assumed based on Schneider and Trauger, 1996 and Brown, 1994
† assumed based on Bagenal, 1994

Table 1.1: Assumed plasma conditions at the time of the Galileo J0 encounter

are equal roughly 2 R_J up the flux tube, where they have a value of approximately 80 kHz, $f_{pe} \sim 0.5 f_{ce}$ when $f_{ce} \sim 100$ kHz. The cyclotron maser instability, a probable source of the decametric emissions, becomes inefficient when $f_{pe} > 0.4 f_{ce}$ and Jupiter's decametric emissions are observed at frequencies of a few hundred kHz and higher [Zarka, 1992].

The Galileo spacecraft has also made direct observations of the Io plasma torus [Gurnett et al., 1996; Frank et al., 1996; Crary et al., 1998], but I only used the Galileo observations for my analysis of ion cyclotron waves in the vicinity of Io. Although Galileo observed a torus which was roughly a factor of two denser than what Voyager 1 encountered [Gurnett et al., 1996], and with a higher oxygen abundance and charge state [Crary et al., 1998], it is unclear which (if either) of these encounters correspond to typical conditions. Estimates of the proton abundances during the Galileo encounter are not yet available. Since the proton density is a major issue for the high latitude processes I describe in chapters 2 and 3, I used the Voyager-era density, composition and temperatures. I wish to use typical, or at least likely, condition in examining the high-latitude phenomena, since I compare my models to Earth-based observations made over a wide period of time. In the case of the ion cyclotron waves observed by Galileo during the Io encounter (chapter 4) I used the Galileo plasma conditions. Unlike Earth-based observations, the local plasma conditions were simultaneously measured at the time of the ion cyclotron observations. Following the analysis of Galileo PLS data by Crary et al., 1998, I assumed that the thermal plasma at this time was isotropic, and use the ion temperatures and abundances given in table 1. These values represent the center of the error bars reported by Crary et al., 1998, for the 1728/1729 SCET PLS spectra, the last set of spectrum taken by the instrument before the Io encounter period. This should accurately describe the upstream conditions during the encounter.

1.7 Overview of thesis

Although Io is a strong source of magnetospheric currents, and Io generates enough power to account for the high-latitude phenomena, there is a gap in our understanding of the interaction. The radio emissions and auroral emissions call for energetic electron beams, not Alfvénic or steady currents. In addition, an Alfvénic disturbance cannot directly transmit a significant amount of energy to the high latitudes: It is strongly reflected by decreasing density and increasing magnetic field strength as it propagates out of the Io plasma torus. In chapters two and three of this thesis, I investigate how the Io interaction reaches the high latitudes and results in particle acceleration. Although uncertainties in the character of the interaction near Io and in the plasma conditions at high latitudes limit my analysis to order of magnitude estimates of electron energies and fluxes, the results are consistent with radio and auroral observations.

In chapter two, I consider an Alfvénic disturbance generated by Io. Although the interaction is not purely Alfvénic , this may be considered either a limiting case, or an approximation of the initial, Alfvénic transient. I derive an analytic description of propagation and reflection if the Io-genic wave. I then apply a non-ideal MHD model, which includes the effects of finite electron inertia and parallel electric fields. I use this and numerical integrations along the Io flux tube to estimate accelerated electron energies and fluxes, both of the electrons which would precipitate and produce auroral features, and of those which would mirror and produce radio emissions. These results are in reasonable agreement with the power required by the main auroral spot and with observations of the S-burst component of the decametric radio emissions. Large parts of this chapter have been published as "On the generation of electron beams by Io", F. J. Crary, *J. Geophys. Res.*,*102*, 37, 1997.

In chapter three, I shift to the implications of a steady current system, which could be present on the down-stream side of Io and which would be present along Io's wake. After considering and dismissing potential drops in the collisional ionosphere, I have examined the effects of double layers as a source of electron acceleration. I show that Birkeland currents, implied by the re-acceleration of Io's wake, are strong and would drive Buneman instabilities in the mid- and high latitudes of the wake flux tube. This instability occurs when the average electron velocity exceeds the thermal velocity. A likely result of this instability is the

formation of double layers. Such double layers accelerate electrons and reduce the potential across Jupiter's ionosphere (thus lengthening the wake/auroral arc.) My estimates of the power and energy of the accelerated electrons are consistent with the Galileo observations of electron beams in Io's wake as well as the estimated power needed to generate Jupiter's decametric radio emissions through the cyclotron maser instability.

In some sense, Io determines the character of its own interaction. The interaction depends on the properties of the Io plasma torus, the medium with which Io interacts. Those properties are determined by Io, since it is the source of that plasma. Ionizations near Io also affect the interaction by heating the plasma and slowing the flow. The Galileo observations of strong ion cyclotron waves near Io provide information on the ionization rates near Io and the nature of Io's neutral exosphere.

In chapter four, I investigate the implications of these waves. First I discuss the limits of quasi-linear diffusion and the reasons why it cannot be applied to this problem. I have used the absence of waves near the S^+ and O^+ cyclotron frequencies to place upper limits on the density of pick-up S^+ and O^+ ions near Io. To do this, I used a warm plasma dispersion analysis to establish a marginal stability criteria, as a function of the pick-up ion's parallel temperature and anisotropy. I numerically calculated the scattering of test particles by the observed wave spectrum, to establish likely values for the ions' parallel temperature and anisotropy. Finally, I have combined these limits on pick-up S^+ and O^+, the estimates of pick-up SO_2^+ density calculated by Huddleston et al. [1997], and the production rates expected for the upsteam conditions observed by Galileo. The final result is an estimate of the column density of SO_2 in Io's exosphere, and upper limits on the column densities of neutral S and O. The SO_2 value is within a factor of a few of the model predictions of Scherb and Smythe [1993], and implies that, based on the observations and analysis of Schneider [1988], the Na abundance in Io's exosphere is approximately 1% at the exobase, increasing to 2% farther from Io.

Chapter 2

Electron acceleration by Alfvén waves

2.1 Introduction

In this chapter, I will describe a model of particle acceleration by an Io-genic. Alfvénic disturbance. Based on the Galileo observations [Kivelson et al., 1996; Frank et al., 1996; Gurnett et al., 1996; Williams et al., 1996], the Io interaction is more complex than purely Alfvénic model (e.g. as described by Goertz [1980] or by Neubauer [1980].) However, my model of particle acceleration concerns the initial disturbance, which would be Alfvénic . The subsequent evolution of the Io interaction, into a steady current loop, would not affect particle acceleration by the initial transient. I will describe a model of particle acceleration by a steady interaction in chapter 3.

When Io crosses a magnetic field line, and starts driving field-aligned currents, the initial transient takes the form of an Alfvénic structure. As this disturbance propagates away from Io, it is reflected by the changing medium of the Io torus. Between Io and Jupiter's ionosphere, the Alfvén velocity increases by approximately a factor of 500. Previous analysis [Goertz, 1980; Wright, 1987] considered the propagation of a plane wave with a frequency of $v_{flow}/R_{Io} \sim 0.03Hz$ and found that a large fraction of the wave's energy would be reflected before reaching Jupiter. This was not considered a serious problem, since the power of the disturbance, $\sim 2x10^{12}W$ based on Voyager data [Acuña et al., 1981] was far larger than the power required to account for the decametric radio emissions, $\sim 5x10^{9}W$. More recently observations, of an Io-related auroral spot, make transmission efficiency a critical issue. With an estimated power of order 10^{11} W, a significant fraction of the initial power must reach the high latitudes in some form. In addition, both the auroral spot and the cyclotron maser instability, the likely source of the decametric emissions, require energetic particles. Currents and Alfvén waves are not a plausible direct cause of either phenomenon. Both the propagation problem and the need for energetic particles implies that, by some process, the energy in the Alfvén wave is converted to energy in the form of energetic particles, which could freely travel to the high latitudes.

I apply a model of Alfvén wave propagation, non-ideal effects to produce an Alfvén wave with a finite, parallel electric field, and Fermi acceleration of electrons. The propagation model differs from previous ones [Goertz, 1980; Wright, 1987], in that I use a Laplace transform and consider a spectrum of waves produced by an impulsive disturbance. Although the

character of this disturbance is approximate, I believe this is a more realistic model than assuming a monochromatic wave. My use of non-ideal Alfvén waves and electron acceleration is similar to models applied to the terrestrial magnetosphere [Kletzing, 1994; Thompson and Lysak, 1996]. I find that this process can generate a beam of electrons with a power of order 10^{11} W and particle energies in the few to hundreds of keV range. This power and energy range is consistent with most observations of the primary auroral spot. In addition, a small fraction of these electrons would mirror. These electrons would have energies of a few keV and represent a power of up to a few times 10^9 W. The higher power levels require including a non-thermal tail on the electron distribution. This power is inadequate to generate the bulk of the decametric emissions, but is sufficient to account for the S-burst decametric emissions. The energy, transient character and loss-cone distribution predicted by model are in good agreement with recent S-burst observations [Zarka et al., 1996; Carr et al., 1997].

2.2 Estimates of the Alfvénic Disturbance

The model of wave propagation and particle interactions discussed in this chapter will require some estimates of the magnitude and form of the Alfvénic disturbance. Alfvén waves are not the only plasma waves generated by Io. In fact, it is impossible for such a disturbance to produce only Alfvén waves [Wright and Schwartz, 1990]. To satisfy conservation of mass, momentum, and energy, the disturbance must also produce fast and slow magnetosonic waves. However, estimates of the other waves produced [Barnett and Olbert, 1986; Linker et al., 1988] suggest that over 70% of the energy generated goes into Alfvén rather than magnetosonic waves. In addition, the magnetosonic mode propagates isotropically, and would not produce any high-latitude effects localized to the Io flux tube. Therefore, for the purposes of the present model, only the Alfvén wave will be considered.

Some analytic descriptions of the Io interaction have been derived [Neubauer, 1980; Goertz, 1980]. However, they only describe the disturbance outside of the Io flux tube (i.e. where parallel current flows may be considered negligibly small) and assume a simple form of the disturbance within the Io flux tube. A full description of the disturbance is not analytically tractable. Instead, it has been modeled by numerical simulations [Linker et al., 1988; Wolf-Gladrow et al., 1987]. For the purposes of this paper, the results of these

numerical models will be used.

The important parameters are the parallel current density, the area across which this current flows and typical scale lengths, perpendicular to the background magnetic field. None of these were measured by the Voyager spacecraft. The magnetometry data show a dipolar magnetic moment of 8.1 A m [Acuña et al., 1981]. But this is an integrated average of current density, weighted by radius. From the measurements, it is not possible to separate radial structure from the magnitude of current density.

The electromagnetic model of *Wolf-Gladrow et al.* [1987] found that parallel currents flow in a annular region between 0.5 and 1 R_{Io} above the surface and that the peak parallel current was $7 \cdot 10^{-7} \frac{A}{m^2}$. While this model was a self-consistent description of the electromagnetic forces, it neglected the inertia of the flowing plasma and its thermal pressure. In essence, the magnetic Reynolds number and plasma β were assumed to be arbitrarily low.

Linker et al., 1988, on the other hand, used a three-dimensional simulation to solve the MHD equations, including realistic parameters for thermal pressure, etc. although their simulation had lower resolution than the model of Wolf-Gladrow et al. Linker et al. found that the parallel current spread over the entire Io flux tube, a circular region roughly 1.2 R_{Io} in radius.

The differences in these models result from different assumptions about the plasma parameters, the nature of Io's ionosphere and their choice of boundary conditions. However, they do have certain common features that may be considered a reasonable description of the Alfvénic disturbance. Both find that parallel current flows in a region which is between one and two R_{Io} from edge to edge and of order 0.5 R_{Io} thick, surrounding the Io flux tube. The magnetic field perturbation quickly goes from zero to approximately 20% of the background field, in this small region of current flow. Within the perturbed Io flux tube, or "Alfvén wings," the perturbation is essentially uniform. These order of magnitude estimates are sufficient for the present work. In addition, an average, parallel current density of 5×10^{-7} A/m^2 will be assumed.

2.3 Current-Carrying Alfvén Waves

Now that the scale and amplitude of the Alfvénic disturbance have been estimated, the characteristics and propagation of that Alfvén wave packet can be examined. To accurately include the effects of low plasma densities and decreasing parallel conductivity, as the wave packet propagates out of the plasma torus, finite electron inertia must be included. This finite inertia results in non-zero, parallel electric fields. Previous models of the Io interactions have neglected this effect. The modified form of the Alfvén wave can be derived in the traditional manner; in cylindrical coordinates with the background magnetic field, B_0 is parallel to the z axis, and $r = 0$ being the center of the flux tube, all wave variables are assumed to be of the form

$$f(r, \phi, z, t) = f_0(r) e^{i(m\phi + kz - \omega t)}$$

and small compared to the background quantities so that only terms which are linear in the perturbation quantities need be retained, and the resulting equations are solved algebraicly. In the case of the Io-generated disturbance, the wave perturbations are roughly 20% of the background quantities, and this linear approximation is only accurate to of order $0.2^2 \sim 5\%$.

The conductivity of oscillations in a cold plasma can be derived from the Fourier transformed equations in Cartesian coordinates. Assuming that the background magnetic field B is in the $\hat{z}$ direction and unperturbed (or that any perturbations are sufficiently small), the velocity of a particle of mass m_α and charge q_α, is

$$\begin{aligned} v_{x,\alpha} &= \left(\tfrac{iq_\alpha}{m_\alpha \omega} E_x - \tfrac{q_\alpha^2 B}{\omega^2 m_\alpha^2} E_y \right) / \left(1 - q_\alpha^2 B^2 / m_\alpha^2 \omega^2 \right) \\ v_{y,\alpha} &= \left(\tfrac{iq_\alpha}{m_\alpha \omega} E_y + \tfrac{q_\alpha^2 B}{\omega^2 m_\alpha^2} E_x \right) / \left(1 - q_\alpha^2 B^2 / m_\alpha^2 \omega^2 \right) \\ v_{z,\alpha} &= \tfrac{iq_\alpha}{m_\alpha \omega} E_z \end{aligned} \tag{2.1}$$

In the limit of $\omega^2 \ll q_\alpha^2 B^2 / m_\alpha^2 = \Omega_{c,\alpha} \ \forall \alpha$, this becomes

$$\begin{aligned} v_x &= -\tfrac{i m_\alpha \omega}{q_\alpha B^2} E_x + \tfrac{1}{B} E_y \\ v_y &= -\tfrac{i m_\alpha \omega}{q_\alpha B^2} E_y - \tfrac{1}{B} E_x \\ v_z &= \tfrac{iq_\alpha}{m_\alpha \omega} E_z \end{aligned} \tag{2.2}$$

The current density is

$$\vec{J} = \sum_{\alpha} n_\alpha q_\alpha \vec{v_\alpha} \tag{2.3}$$

Applying the requirement of charge neutrality, $\sum_\alpha n_\alpha q_\alpha = 0$ and the definition of mass density, $\sum_\alpha n_\alpha m_\alpha = \rho$, the conductivity relations are

$$\begin{aligned} \mu_0\sigma_\perp &= \tfrac{-i\mu_0\rho\omega}{B_0^2} = -i\tfrac{\omega}{V_A^2} \\ \mu_0\sigma_\parallel &= \tfrac{i\mu_0 e^2 n_e}{m_e\omega} = i\tfrac{\omega_p^2}{c^2\omega} \end{aligned} \tag{2.4}$$

In cylindrical coordinates, and after Fourier transforming with respect to t, z, and ϕ and setting $b_z = 0$ (to select the parallel-propagating, Alfvén wave), Maxwell's equations take the form

$$\begin{aligned} \omega b_r &= \tfrac{m}{r}E_z - kE_\phi \\ \omega b_\phi &= kE_r + i\tfrac{\partial}{\partial r}E_z \\ \tfrac{\partial}{\partial r}(rE_\phi) &= imE_r \\ \left(\tfrac{-i\omega}{c^2} + \mu_0\sigma_\perp\right)E_r &= -ikb_\phi \\ \left(\tfrac{-i\omega}{c^2} + \mu_0\sigma_\perp\right)E_\phi &= ikb_r \\ \left(\tfrac{-i\omega}{c^2} + \mu_0\sigma_\parallel\right)E_z &= \tfrac{1}{r}\left[\tfrac{\partial}{\partial r}(rb_\phi) - imb_r\right] \end{aligned} \tag{2.5}$$

Inserting the conductivity leads to the relations,

$$b_\phi = \frac{i}{m}\frac{\partial}{\partial r}(rb_r) \tag{2.6}$$

$$E_r = \frac{ikV_A^2}{m\omega\left(1+\frac{V_A^2}{c^2}\right)}\frac{\partial}{\partial r}(rb_r) \tag{2.7}$$

$$E_\phi = -\frac{kV_A^2}{\omega\left(1+\frac{V_A^2}{c^2}\right)}b_r \tag{2.8}$$

$$E_z = \frac{r\omega}{m\left(1+\frac{V_A^2}{c^2}\right)}\left(1+\frac{V_A^2}{c^2}-\frac{k^2V_A^2}{\omega^2}\right)b_r \tag{2.9}$$

and the differential equation

$$r^2\frac{\partial^2}{\partial r^2}b_r + 3r\frac{\partial}{\partial r}b_r + \left[\lambda^2r^2 - (m^2-1)\right]b_r = 0 \tag{2.10}$$

$$\lambda^2 = \frac{\omega^2 - \omega_p^2}{c^2\left(1 + \frac{V_A^2}{c^2}\right)}\left(1 + \frac{V_A^2}{c^2} - \frac{k^2 V_A^2}{\omega^2}\right) \tag{2.11}$$

This equation has solutions of the form

$$b_r = b_{r,0}\frac{1}{\lambda r}J_m(\lambda r) \tag{2.12}$$

$$b_r = b_{r,0}\frac{1}{\lambda r}I_m(i\lambda r) \tag{2.13}$$

where J_m and I_m are the Bessel functions and modified Bessel functions of the first kind. Y_m and K_m are also solutions, but will not be considered since $Y_m, K_m \to \infty$ as $r \to 0$. The I_m will also be ignored, since $I_m \to \infty$ as $r \to \infty$.

In the low-frequency, non-relativistic limit, the dispersion relation simplifies to

$$\frac{k^2}{\omega^2} = \frac{1}{V_A^2}\left(1 \pm \lambda^2 L^2\right) \tag{2.14}$$

$$L = \frac{c}{\omega_p} \tag{2.15}$$

Finite electron inertia also adds a cutoff for the I_m and K_m solutions,

$$n = \frac{ck}{\omega} \to 0 \tag{2.16}$$

as

$$\lambda \to \frac{\omega_p}{c} \tag{2.17}$$

For shorter perpendicular length scales, the I_m and K_m waves become evanescent, while the J_m and Y_m waves continue to propagate with a dispersive group velocity. In the Io plasma torus and along the associated flux tubes, $50\ m < L < 10^4\ m$. The effect of finite electron inertia becomes important as L becomes large, that is, when n_e and hence ω_p becomes small.

In addition to the effects of finite electron inertia, the Alfvén wave may also be non-ideal due to kinetic effects. This may be neglected when the $\lambda r_L \ll 1$, where r_L is the gyroradius of thermal ions. The relative importance of electron inertia and kinetic effects is determined by $L/r_L = c\Omega_{c,i}/\omega_{p,e}v_{t,i} \approx \sqrt{m_e/M_i\beta}$, where β is the ratio of thermal to magnetic pressure. Electron inertia is dominant when $\beta < m_e/M_i$. While the Io torus is a

low β plasma, it is not quite that low in the vicinity of Io. However, near Io, $L \sim 0.1$ km, $r_L \sim 2.5$ km, and $\lambda < 10R_{Io}^{-1} = 0.005km^{-1}$ for the Io-genic Alfvén wave. So $\lambda r_L, \lambda L \ll 1$ near Io. Although the non-ideal effects near Io are kinetic rather than electron inertial, this may be neglected since all non-ideal effects are extremely weak in this region. At higher latitudes, as the wave propagates out of the torus and into stronger magnetic fields, L increases, while r_L decreases. As a result, the kinetic effects never become significant while the electron inertial effects are important in the mid- and high latitudes.

Of particular interest to the Io-Jupiter problem are the energy (Poynting) flux, electric potential, current and particle velocities along field lines. These are related to the magnetic field perturbation by

$$F = \frac{V_A b_0^2}{\mu_0}\left(1 + \frac{V_A^2}{2c^2} + \frac{L^2\lambda^2}{2m^2}\right) \tag{2.18}$$

$$J_z = \frac{-i\lambda}{\mu_0 m} b_0 J_m(\lambda r) e^{i(kz-\omega t+m\phi)} \tag{2.19}$$

$$v_{ez} = \frac{i\lambda}{\mu_0 e m n_e} b_0 J_m(\lambda r) e^{i(kz-\omega t+m\phi)} \tag{2.20}$$

The electric potential in the wave's reference frame, may be directly calculated from the electrons' wave-induced velocity by applying conservation of energy

$$\begin{aligned} -e\Phi + \frac{m_e}{2}(v_{ez} - V_A)^2 &= \frac{m_e}{2}V_A^2 \\ \Phi &\approx -\tfrac{m_e}{e} V_A v_{ez} \\ &\approx \tfrac{\epsilon_0}{-\omega_p^2} V_A J_z \\ &\approx \tfrac{\epsilon_0}{-\omega_p^2} V_A k_\perp b_0 \end{aligned} \tag{2.21}$$

Since the medium is assumed to be uniform perpendicular to the magnetic field, $k_\perp$ only depends on the width of the flux tube. b_0 is determined by a propagation model, described in the following section. The magnitude of the potential does not depend on the details of the non-ideal Alfvén wave. Whether the wave is kinetic or inertial, equation 2.21 gives the associated potential. The specifics of the non-ideal wave determine $k_\parallel$, V_A and $E_\parallel$, but the effect on V_A is weak in the case of the Io-genic Alfvén wave and the exact form of $k_\parallel$ and $E_\parallel$ are not required by the particle acceleration mode presented below. Therefore, the results

given in this section are general, and apply whether or not the Alfvén wave is kinetic or inertial in the vicinity of Io. On the basis of the model of wave propagation, described in the following section, the Io-generated wave packet would have a potential ranging from $1 \cdot 10^{-3}$ to 75 V.

The properties which affect wave propagation are the phase and group velocities and the energy flux. If these quantities are the same for both the current-carrying and the plane Alfvén wave, then the existing work on (plane) Alfvén wave propagation may be applied. From (2.14), the phase and group velocities are nearly identical to the plane wave so long as $\lambda L \ll 1$. The perpendicular scale length of the disturbance seems to be greater than 0.1 R_{Io}, near Io, or $\sim$ 200 km when mapped to the foot of Io's flux tube. Therefore, $\lambda L \ll 1$ for $\omega_p \gg 1500$ s^{-1} or $n_e \gg 0.001$ cm^{-3}. This is the case in all parts of the Io flux tube. Equation (2.18) imposes the same requirement for the energy flux. As a result, the only effect of finite electron inertia is to produce a parallel electric field. A generic propagation model, which considers neither finite electron inertia nor the geometry of the current flow, will still accurately describe the wave's propagation.

2.4 Propagation in a Nonuniform Medium

The gradual change in the density and magnetic fields near Io suggests the WKB approximation could be applied. However, this formalism is only valid if the length scale of the medium, $L \equiv V_A(z)/(\partial V_A(z)/\partial z)$, is large compared to the wavelength, or, expressed in terms of frequency

$$\frac{\omega}{\frac{\partial V_A}{\partial z}} \gg 1 \tag{2.22}$$

As Figure 2.1 shows, the WKB approximation is valid close to Io but breaks down within 2 R_J of the centrifugal equator. As a result, some other description of wave propagation is required. *Goertz* [1980] derived such a solution by assuming that an Alfvén velocity of the form

$$\frac{1}{V_A^2} = \frac{1}{V_{A\infty}^2} + \left(\frac{1}{V_{A0}^2} - \frac{1}{V_{A\infty}^2}\right) e^{-2z/z_0} \tag{2.23}$$

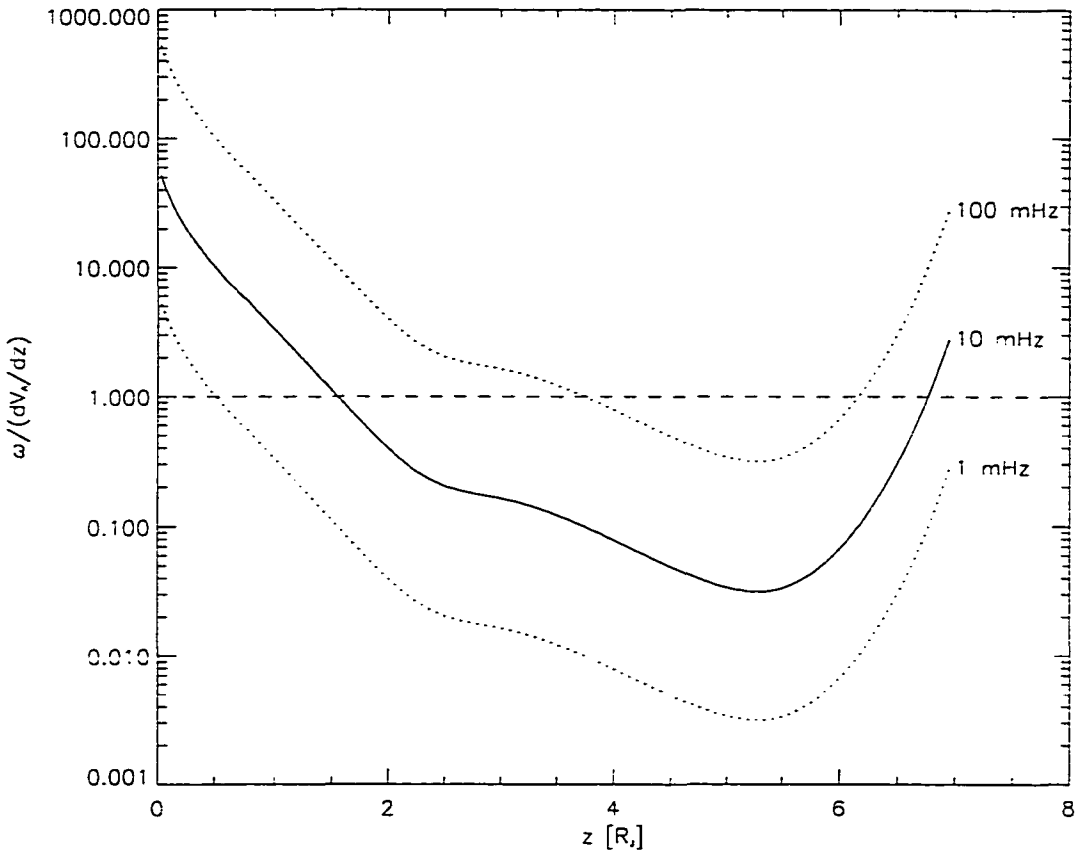


Figure 2.1: The value of $\omega / \frac{\partial V_A}{\partial z}$ as a function of latitude, for three frequencies. The WKB approximation is valid if this quantity is much greater than 1.

However, his solution only allowed changes in ρ and required the magnetic field to be constant. Following *Ferraro and Plumpton* [1957], Goertz solved the wave equation,

$$V_A^2(z)\frac{\partial^2 v}{\partial z^2} = \frac{\partial^2 v}{\partial t^2} \tag{2.24}$$

While this is correct for the approximation used by both Ferraro and Goertz, it is only valid if B_0, the background magnetic field, is constant and the density, ρ is a function of z. If B_0 varies with z, then the wave equation takes the form,

$$\frac{\partial}{\partial z}\left(V_A^2(z)\frac{\partial}{\partial z}b\right) = \frac{\partial^2}{\partial t^2}b \tag{2.25}$$

This neglects the displacement current, but that term is only significant when V_A is of order c. As can be seen in Figure 2.2, this does not occur until $z > 5.5R_J$. As will be shown below, the wave is almost completely reflected before reaching this region.

Following the technique of Goertz and of Ferraro, the Alfvén velocity will be approximated by,

$$V_A = V_{A0}e^{z/z_0} \tag{2.26}$$

Figure 2.2 shows the Alfvén velocity along the Io flux tube using the diffusive equilibrium model [Bagenal, 1994]. The exponential approximation is reasonably accurate for most of the flux tube. It is inaccurate near the centrifugal equator and very close to Jupiter. Neither case will significantly alter the results. Near the centrifugal equator, the WKB approximation is valid. So the wave may be assumed to propagate from Io in a WKB manner, until it has traveled roughly 1.5 R_J. Beyond that point, it will propagate as according to the model presented below. This model will be valid until the wave reaches the high latitudes where the Alfvén wave becomes relativistic.

In addition, the exponential approximation may allow z_0 to vary. Even if the Alfvén velocity is not truly exponential, z_0 may be defined as

$$z_0 = z/\ln\frac{V_A}{V_{A0}} \tag{2.27}$$

Since $\partial/\partial z_0$ is not zero, this assumption will cause errors from both local variations in z_0 and

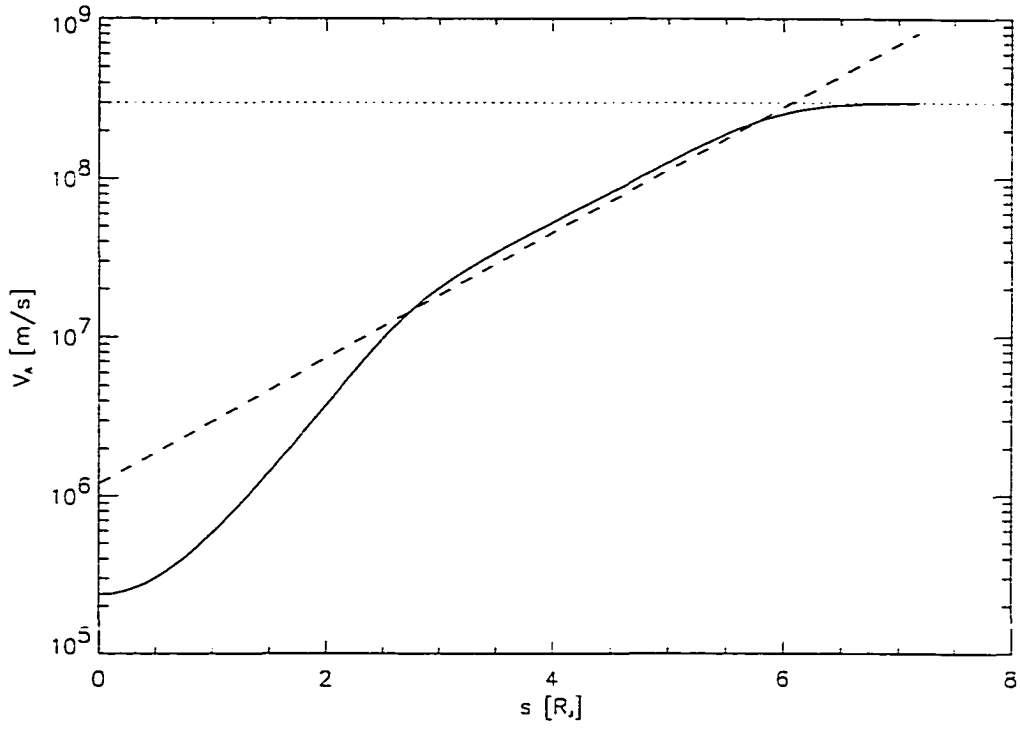


Figure 2.2: Alfvén velocity according to the diffusive equilibrium models (solid line) and according to an exponential approximation (dashed line). For most of the mid-latitude region, the velocity is approximately exponential with distance along the field line. The speed of light is shown as a dotted line.

cumulative differences between the actual and modeled wave propagation. In the equations that follow, such a variable z_0 would produce additional, local terms of order $\partial/\partial z_0$. Figure 2.3 shows z_0 for an Alfvén velocity along the Io flux tube. This is based on the density model of *Bagenal* [1994] and the O4 octapole magnetic field model. Figure 2.4 shows $\partial z_0/\partial z$. Since $|\partial z_0/\partial|z \ll 1$ for most of the flux tube, a model based on an exponential increase in the Alfvén velocity may accurately be applied. The cumulative errors are assumed to average to zero, at least to first order. These errors would be integrals of some function of $V_A - V_{A0}e^{z/z_0}$ along the field line. Since the exponential is a good approximation to V_A, such an integral would be zero to first order.

Equation (2.25) can be solved by either Laplace or Fourier transform methods and a change of variables. Previous work [Goertz, 1980; Wright, 1987] used the Fourier method and the assumption of a single, monochromatic wave. The actual disturbance is more likely to be a complex superposition of many different frequencies. While this spectrum is unknown and depends strongly on the details of the Io interaction, the Laplace transform method may provide a more realistic description: It inherently gives the spectrum produced by specified, initial conditions.

Alfvén wave propagation, allowing for a varying magnetic field strength and mass density, is described by

$$\frac{\partial}{\partial z}\left(V_A^2(z)\frac{\partial}{\partial z}b\right) = \frac{\partial^2}{\partial t^2}b \tag{2.28}$$

Using a Laplace transform and the change of variables,

$$x = \frac{sz_0}{V_{A0}}e^{-z/z_0} \tag{2.29}$$

where s is the Laplace transform variable, the wave equation becomes

$$-\frac{1}{x}b(s) + \frac{\partial}{\partial x}\left(\frac{1}{x}\frac{\partial}{\partial x}b(s)\right) = \frac{-1}{sx}\,b(t)|_{t=0} - \frac{1}{s^2x}\,\frac{\partial b(t)}{\partial t}\bigg|_{t=0} \tag{2.30}$$

The homogeneous equation has the solutions

$$b = \begin{cases} ixI_1(x) \\ ixK_1(x) \end{cases} \tag{2.31}$$

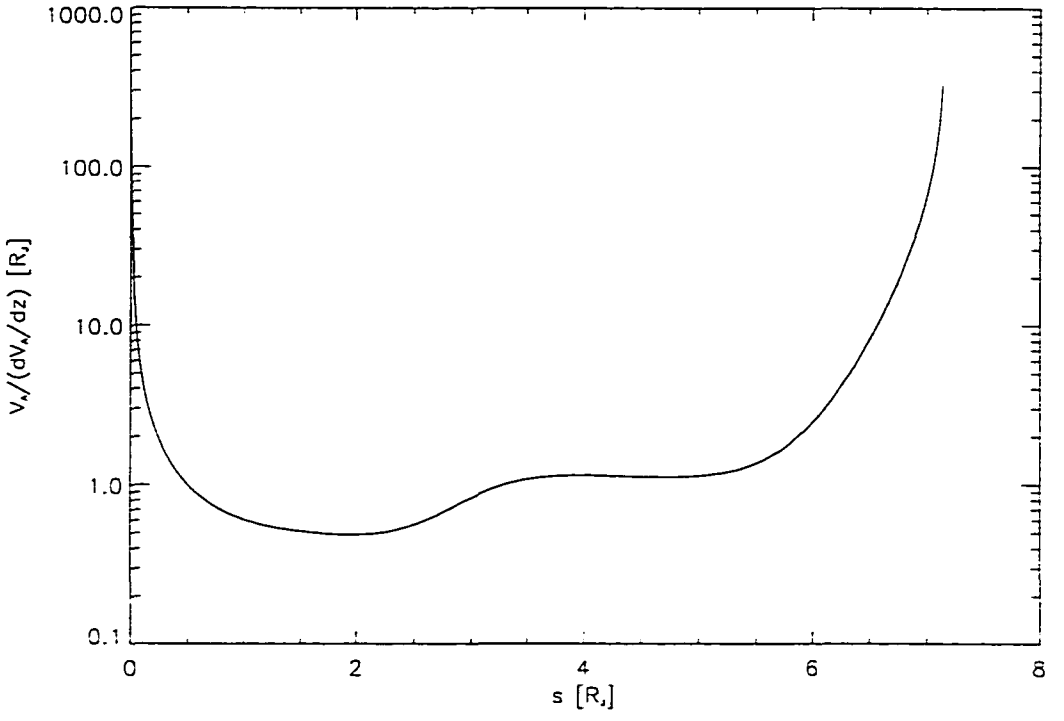


Figure 2.3: Alfvén scale height z_0 calculated from the diffusive equilibrium model. Between 1.5 and 6 R_J, the scale height is nearly constant, and the Alfvén velocity is well-approximated by an exponential.

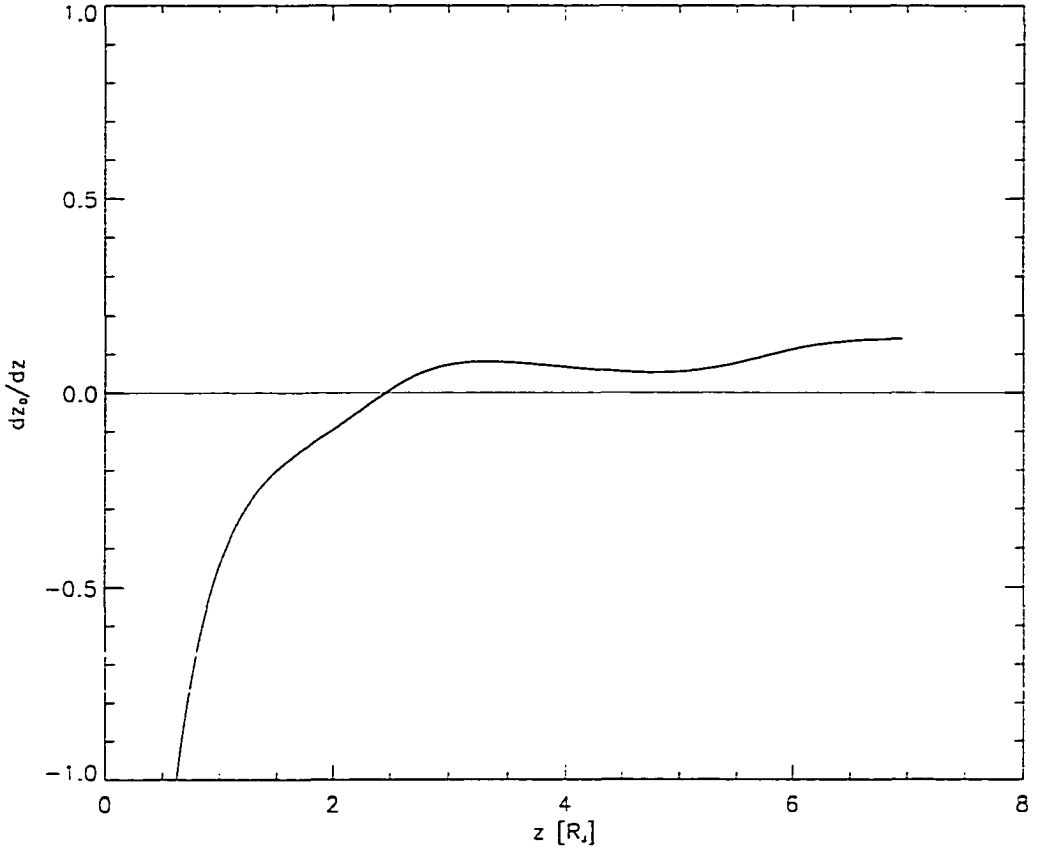


Figure 2.4: Rate of change in Alfvén scale height. If $\partial z_0/\partial z \ll 1$, the results of wave propagation in an exponential medium may be applied with reasonable accuracy.

where I and K are the modified Bessel functions of first order. The inhomogeneous solution can be found by finding the Green's function solution and performing an inverse Laplace transform. As noted earlier, the perturbation within the Alfvén wings is constant; $\partial b/\partial t = 0$ except in a small region around the Io flux tube. For this reason, the $\partial b/\partial t$ forcing term in equation (2.30) will be neglected, and the forcing approximated by

$$b|_{t=0} = b_0 \delta(x - x_0) \tag{2.32}$$

$$\left.\frac{\partial b}{\partial t}\right|_{t=0} = 0 \tag{2.33}$$

where $x_0 = sz_0/V_{A0}$ is the value of x at z_0. For which the Green's function is

$$b(s) = \frac{-b_0}{s} \begin{cases} xK_1(x_0)I_1(x) & 0 < x < x_0 \\ xK_1(x)I_1(x_0) & x_0 < x \end{cases} \tag{2.34}$$

For the Alfvénic disturbance propagating out of the Io torus, the solution for $x < x_0$ ($z > z_0$) is relevant. The solution for b as a function of z may be found by the inverse Laplace transform,

$$b(x,t) = \frac{-ib_0\beta}{2\pi} \int_{c-i\infty}^{c+i\infty} K_1(\alpha s) I_1(\beta s) e^{st} ds \tag{2.35}$$

$$\alpha \equiv \frac{x_0}{s} = \frac{z_0}{V_{A0}} \tag{2.36}$$

$$\beta \equiv \frac{x}{s} = \frac{z_0}{V_{A0}} e^{-z/z_0} \tag{2.37}$$

Physically, α and β represent propagation times. If the Alfvén velocity really increased exponentially, for all z, then the wave would require a time, α, to propagate from $z = z_0$ to $z = \infty$. This is, of course, a mathematical fiction, since the Alfvén velocity would have to similarly go to infinity, rather than its actual, relativistic limit. In any case, the system is closed. At approximately $z = 7\ R_J$, where z is the distance from Io along the magnetic field line, the wave reaches the Jovian ionosphere, and the disturbance can go no farther. This occurs at the time, $t = 0.9996\alpha$. It is a good approximation to consider α the propagation time from $z = z_0$ to Jupiter's ionosphere. β is similarly the propagation time from the point, z, to the Jovian ionosphere. As a result, $\alpha - \beta$ is the propagation time from z_0 to z. The

present model does not consider or attempt to model the properties of the Jovian ionosphere, nor account for its reflection of the Alfvénic disturbance. Therefore only times $0 < t < \alpha$ will be considered.

For $t < \alpha - \beta$, equation (2.35) can be solved as a contour integral closed in the positive real side of the complex plane,

$$b = 0 \qquad t < \alpha - \beta = \frac{z_0}{V_{A0}} \left(1 - e^{-z/z_0}\right) \tag{2.38}$$

This is a causality condition and represents the undisturbed state of the system before the wave packet has had time to propagate to the point, z.

Unfortunately, this technique can not be used to find the solution for times $\alpha - \beta < t < \alpha + \beta$. Solutions for later times, $t > \alpha + \beta$ can be found by a contour integral closed in the negative real half of of the complex plane. However, these solutions are for times long after the disturbance has reached Jupiter's ionosphere. They are therefore neither interesting nor meaningful.

Although the traditional techniques for the inverse Laplace transform do not produce a satisfactory solution, equation (2.35) can be solved by a change of variables and a symmetry condition. Taking $c = 0, s = -ip$, the equation becomes

$$\begin{aligned} b(x,t) &= \tfrac{b_0\beta}{2\pi} \int_{-\infty}^{\infty} K_1(-i\alpha p) I_1(-i\beta p) e^{-ipt} ds \\ &= \tfrac{b_0\beta}{2\pi} \int_{0}^{\infty} K_1(-i\alpha p) I_1(-i\beta p) e^{-ipt} + K_1(i\alpha p) I_1(i\beta p) e^{ipt} ds \end{aligned} \tag{2.39}$$

Noting that

$$\begin{aligned} K_1(ip) &= -\tfrac{\pi}{2} \left[J_1(p) - iY_1(p)\right] \\ K_1(-ip) &= -\tfrac{\pi}{2} \left[J_1(p) + iY_1(p)\right] \\ I_1(ip) &= iJ_1(p) \\ I_1(-ip) &= -iJ_1(p) \end{aligned} \tag{2.40}$$

The integral reduces to

$$b = \frac{-b_0\beta}{2} \left[\int_0^{\infty} dp J_1(\beta p) J_1(\alpha p) \sin(pt) - \int_0^{\infty} dp J_1(\beta p) Y_1(\alpha p) \cos(pt) \right] \tag{2.41}$$

The first integral has tabulated solutions,

$$\int_0^\infty dp J_1(\beta p) J_1(\alpha p) \sin(pt) =$$

$$\begin{cases} 0 & t < \alpha - \beta \\ \frac{1}{2\sqrt{\alpha\beta}} P_{1/2}\left(\frac{\alpha^2+\beta^2-t^2}{2\alpha\beta}\right) & \alpha - \beta < t < \alpha + \beta \\ \frac{-1}{\pi\sqrt{\alpha\beta}} Q_{1/2}\left(\frac{t^2-\alpha^2-\beta^2}{2\alpha\beta}\right) & \alpha + \beta < t \end{cases} \tag{2.42}$$

where $P_{1/2}$ and $Q_{1/2}$ are the Legendre functions of half integral order [Erdélyi et al., 1953]. The second integral in the inverse Laplace transform does not have tabulated solutions. However, a solution may be obtained using the inherent symmetry of the functions (T. J. Bogdan, personal communication, 1994). Consider the system at a time, τ second before the initial disturbance, i.e. at $t = -\tau < 0$. Causality, and the solution given above, require that

$$0 = b(x,t) \quad t < 0 < \alpha - \beta \tag{2.43}$$

$$\begin{aligned} 0 &= b \\ &= \frac{-b_0\beta}{2}\left[\int_0^\infty dp J_1(\beta p) J_1(\alpha p) \sin(-p\tau) - \int_0^\infty dp J_1(\beta p) Y_1(\alpha p) \cos(-p\tau)\right] \\ &= -\frac{-b_0\beta}{2}\left[\int_0^\infty dp J_1(\beta p) J_1(\alpha p) \sin(p\tau) - \int_0^\infty dp J_1(\beta p) Y_1(\alpha p) \cos(p\tau)\right] \end{aligned} \tag{2.44}$$

$$-\int_0^\infty dp J_1(\beta p) J_1(\alpha p) sin(pt) = \int_0^\infty dp J_1(\beta p) Y_1(\alpha p) cos(p\tau) \tag{2.45}$$

Therefore the amplitude of the magnetic field disturbance is

$$b(z,t) = b_0 \begin{cases} 0 & t < t_0 \\ e^{\frac{-z}{2z_0}} P_{1/2}\left[\cosh\left(\frac{z}{z_0}\right) - \frac{t^2}{2t_1^2} e^{\frac{z}{z_0}}\right] & t_0 \le t < t_1 \end{cases}, \tag{2.46}$$

$$t_0 = \frac{z_0}{V_{A0}}\left(1 - e^{z/z_0}\right)$$

$$t_1 = \frac{z_0}{V_{A0}}$$

As noted earlier, the disturbance reaches the Jovian ionosphere at $t \sim z_0/V_{A0}$, and events after that time are beyond the scope of this model.

The Legendre function represents the phase and shape of the wave packet, and defines the phase fronts of the disturbance, while the exponential term describes the decrease in amplitude as the wave packet is reflected. It is convenient to note that $e^{-z/2z_0} = \sqrt{V_{A0}/V_A(z)}$.

Since $V_A/c \ll 1$ and $L\lambda \ll 1$, equation 2.18 gives an energy flux proportional to $V_A b^2$. The total power transmitted is this flux times the cross-sectional area of the flux tube, proportional to $1/B$. As a result, the transmission coefficient for the plasma is B_0/B, roughly 0.01 between Io and the Jovian ionosphere. This solution is similar to those found by other models [Goertz, 1980; Wright, 1987; Wright and Schwartz, 1989] and leads to the following problem: Before reaching the high latitudes, the Alfvén velocity increases by a factor of roughly 500. This implies that approximately 99% of the Alfvén wave's energy is reflected and never reaches the high latitudes. Despite this, energetic, Io-related phenomena are observed at high latitudes. The power produced by the Io interaction must be transmitted all the way to Jupiter's ionosphere, but it can not be transmitted by an Alfvén wave or wave packet.

2.5 Acceleration of an Electron Beam

The Alfvén wave's parallel electric field will accelerate particles. An early paper on the Io-Jupiter current flow [Goldreich and Lynden-Bell, 1969] asserted that the current itself must be carried by a beam of keV electrons but offered no clear mechanism for accelerating these particles. More detailed work [Smith and Goertz, 1978] noted that a steady state current system would develop an electrostatic double layer (i.e. a region of strong, stationary, parallel electric fields) which would be a plausible mechanism for accelerating the electrons. However, these efforts were time-independent models based on the pre-Voyager assumption of steady-state current flow.

Other work, concerning the Earth's magnetosphere, has considered an Alfvén wave with a parallel electric field, either as a result of kinetic [Goertz, 1984] or geometric effects [Goertz and Boswell, 1979]. Such a wave would act as an "electrostatic shock" and accelerate particles. This theory has been used to accurately model the source of energetic particles responsible for the terrestrial aurora [Hui and Seyler, 1992; Kletzing, 1994; Thompson and

Lysak, 1996] but has never been applied to the Io-Jupiter current flow. The process also becomes more complex because of the non-uniform medium.

The basic mechanism considered here is Fermi acceleration, essentially a collision between a stationary (or slowly moving) particle and a moving barrier, in this case, an electric potential. The result is clear when considered in the wave's rest frame. Initially, the particle has a velocity

$$v_i = v_0 - V_A \tag{2.47}$$

where V_A is the wave velocity and v_0 is the particle's velocity in the rest frame. Upon colliding with the potential barrier, the particle reflects and has a velocity

$$v_f = -v_i = V_A - v_0 \tag{2.48}$$

Transforming back into the stationary frame,

$$v_f = 2V_A - v_0 \tag{2.49}$$

The collision accelerates the particle to nearly twice the phase velocity of the wave and increases its energy by

$$\Delta E = 2mV_A(V_A - v_0) \tag{2.50}$$

On the surface, it appears that accelerated protons would be more energetic and significant than electrons. However, the wave will not accelerate every particle it encounters. First, the particle and wave must be moving towards each other:

$$v_i \leq 0 \Rightarrow v_0 \leq v_A \tag{2.51}$$

Second, to accelerate a particle, the wave's electric potential must be large enough to cause a reflection. That is,

$$v_0 \geq V_A - \sqrt{\frac{2e}{m}\Phi} \tag{2.52}$$

The other particles are simply slightly accelerated and then decelerated back to their initial velocity as the wave passes. This is the electron motion which produces the wave's current.

As a result, this process only affects particles with an initial velocity,

$$V_A - \sqrt{\frac{2e}{m}\Phi} \leq v_0 \leq V_A \tag{2.53}$$

Clearly, this criterion is much narrower for protons than for electrons. In addition, the proton thermal velocity is much lower than the electron thermal velocity (by a factor of 43 if they were at the same temperature, 12.9 times lower for the conditions measured near the Io flux tube). This further reduces the number of protons at or near the Alfvén velocity. The accelerated protons are few enough that they can safely be neglected.

For particles in a Maxwellian distribution, the number of accelerated particles would be,

$$n = n_e \int_{\infty}^{\infty} dv_x \int_{\infty}^{\infty} dv_y \int_{v_1}^{v_2} dv_z f(\vec{v}) \tag{2.54}$$

$$= \frac{1}{2} n_e \left[\mathrm{Erf}(\frac{V_1}{v_t}) - \mathrm{Erf}(\frac{V_2}{v_t}) \right] \tag{2.55}$$

Where $V_1 = V_A - \sqrt{m_e/m} J_z/(en_e)$, $V_2 = V_A$ and v_t is the thermal velocity.

This acceleration process assumes that the wave is propagating at a constant velocity. If, on the other hand, the group velocity is increasing, multiple collisions can occur. A particle approaching the wave with at a relative velocity v_{p0}, will, after the first collision, have a constant velocity,

$$v_{pf} = 2V_A - v_{p0} \tag{2.56}$$

Immediately after the collision, this may be much greater than the wave's group velocity. However, as V_A increases, the wave will overtake the particle and cause another collision. This will occur where

$$\int_{z_i}^{z_i+\Delta z} \frac{dz}{V_A(z)} = \frac{\Delta z}{v_p} \tag{2.57}$$

Following the assumption that $V_A = V_{A0} e^{z/z_0}$, this distance between collisions, Δz, is

$$\frac{\Delta z}{z_0} \frac{1}{1 - e^{-\Delta z/z_0}} = \frac{v_{pf}}{V_A(z_i)} \tag{2.58}$$

While this is not solvable analytically, it is not difficult to obtain a numerical answer. In

general, however, v_{p0} is always close to V_A. Figure 2.5 shows the ratio of V_A and v_{p0} along the Io flux tube, based on the temperatures given by *Bagenal* [1994], the above equations and the wave amplitude calculated in the next section of this paper. The beam electrons have velocity within a few percent of the Alfvén velocity. As a result, equation (2.58) predicts repeated collisions every 10^2 to 10^5 m, with any given particle encountering the wave at least once every 0.5 s. Since these events are frequent and closely spaced, the individual collisions may be neglected and the process treated as a continuous acceleration of the electrons.

Between collisions the particle will have a velocity

$$V_A - \Delta V \leq v_p \leq V_A + \Delta V \tag{2.59}$$

where $\Delta V = v_{pf} - V_A \ll V_A$. These repeated collisions will continue as long as

$$\Delta V \leq \sqrt{\frac{2e}{m}\Phi} = \sqrt{\frac{2}{en_e}V_A J_z} \tag{2.60}$$

Once this condition is violated, the wave and particles will decouple and move along the field lines separately.

2.6 Approximate Electron Flux and Energy

The number of the accelerated electrons may be calculated by integrating equation (2.55) over the distance.

$$\begin{aligned} F &= \int_0^{z_f} ds \frac{B_0}{B(s)} n(s) \\ &= \frac{1}{2} \int_0^{z_f} ds \frac{B_o}{B(s)} n_e(s) \left[\mathrm{Erf}(\tfrac{V_1(s)}{v_t(s)}) - \mathrm{Erf}(\tfrac{V_2(s)}{v_t(s)}) \right] \end{aligned} \tag{2.61}$$

In general, this integral does not have an analytic solution. However, it may be calculated numerically. Models of the Io torus give $n_e(s)$, $v_t(s)$ and V_A as a function of position along the Io flux tube [Acuña et al., 1983; Bagenal, 1994]. This more accurate Alfvén velocity was used, rather than an exponential approximation. Given these models, the number and energy of accelerated electrons were calculated, as well as what fraction of the electrons were

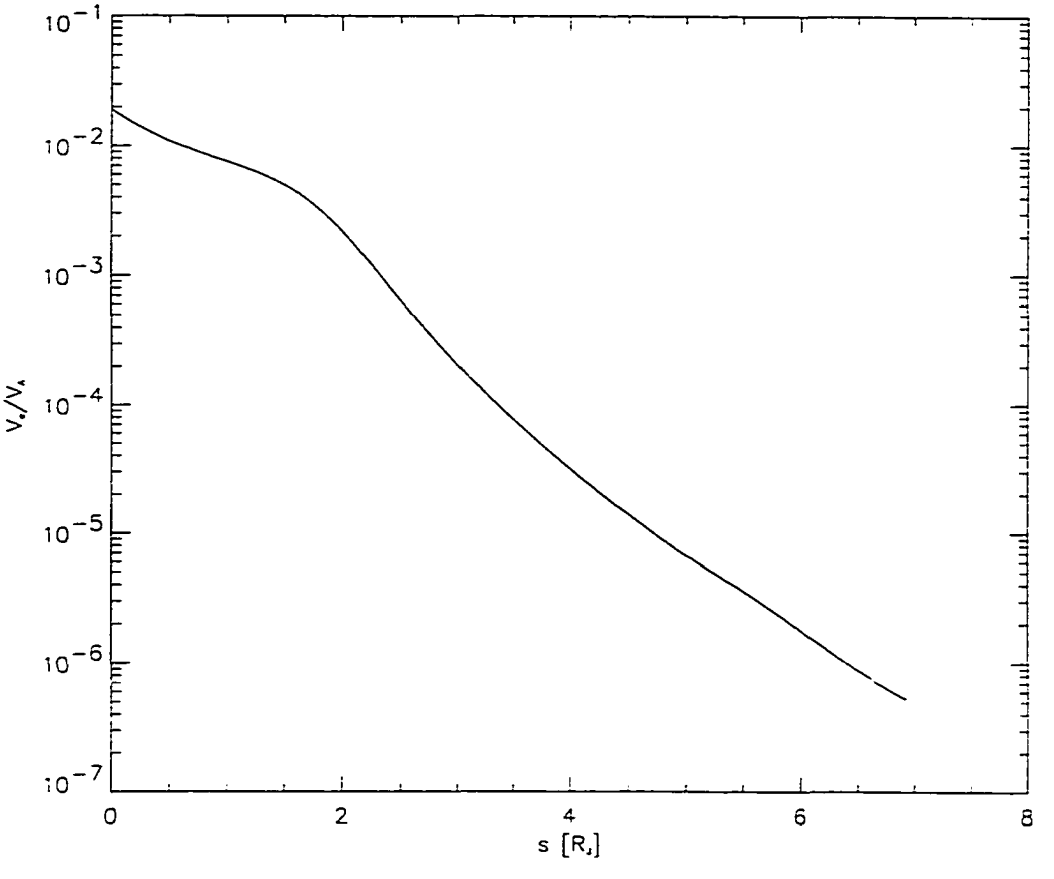


Figure 2.5: Ratio of the maximum relative velocity for reflected electrons, V_e to that of the Alfvén wave, V_A. Since V_e/V_A is always a small number, the reflected electrons will frequently collied with the Alfvén wave.

outside of the loss cone. First, the magnetic field strength, Alfvén velocity and electron density along the flux tube were calculated. This was done at 294 points along the field line, spaced 2x10^6 m (0.028 R_J) apart. These values, an assumed initial amplitude, and the above equations were used to calculate, V_1, V_2 and the number of electrons trapped during the first step, $0 < z < 2\text{x}10^6$ m, and which decouple before reaching 2x10^6 m. The energy of these particles, and the energy they gained from the wave was determined by equations (2.53) and (2.50). On the basis of equation (2.46), the wave amplitude along the second step, $2 < z < 4\text{x}10^6$ m, was calculated. This amplitude was further decreased, based on the energy gained by trapped electrons during the first step. The actual process by which the wave looses energy is beyond the scope of this chapter, but conservation of energy requires that all energy gained by the electrons must be lost by the wave. This amplitude was then used to calculate the number, energy, etc. of electrons which decouple from the wave before reaching 4x10^6 and were trapped either in the first or second step. This, in turn was used to determine the wave amplitude along the third step, and the process was repeated for each subsequent step. The end result is the amplitude of the wave at 249 points along the flux tube and a matrix of 249x249 elements, where element i, j is the number of electrons which were trapped in step i and which decoupled in step j. From these, and the field strength and Alfvén velocity, many interesting values may be calculated (e.g., total power of the electron beam, the fraction of electrons outside the loss cone, etc.)

Figure 2.6 shows the potential along the magnetic field line, caused by the Alfvén wave, for the case of Io at a System III longitude of 112^o. Figure 2.7 shows the integrand of equation 2.61 as a function of s The total flux of electrons reaching Jupiter's ionosphere is approximately 8x10^{24} per second. These electrons have a velocity approximately equal to the Alfvén velocity at the location where the wave and particle decoupled. Figure 2.8 shows the resulting spectrum of precipitating electrons. Figure 2.9 shows the rate at which particles decouple from the wave, as a function of position. The average particle energy is 78 keV and the total power in the electron beam is 9.9x10^{10} W.

Both the spectrum and the decoupling rate have a bimodal nature. The low energy electrons reaching Jupiter have a broad distribution centered near 1 keV, but extending up to over 100 keV. This is a result of electrons which gradually decouple from the wave, as a result of its reflection. Until the wave has traveled roughly 5 R_J ($\sim 42^o$ latitude), reflection from

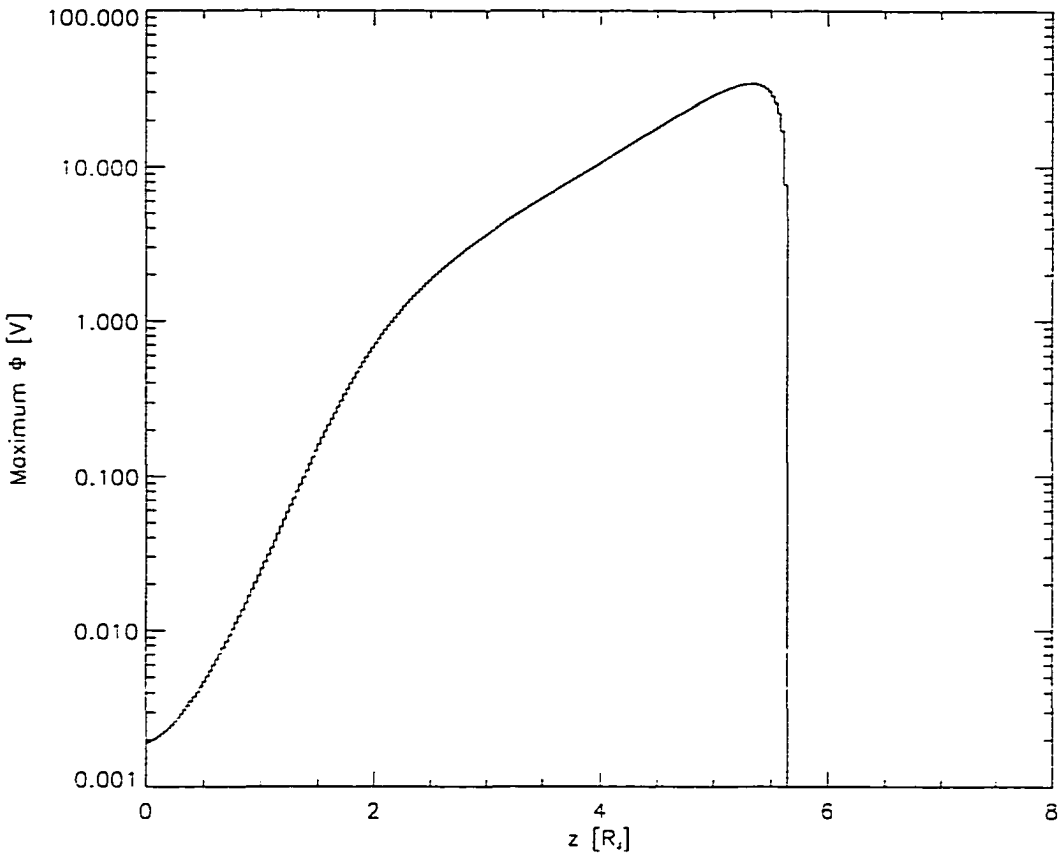


Figure 2.6: The parallel electric potential generated by the Io-genic Alfvén wave because of finite electron inertia.

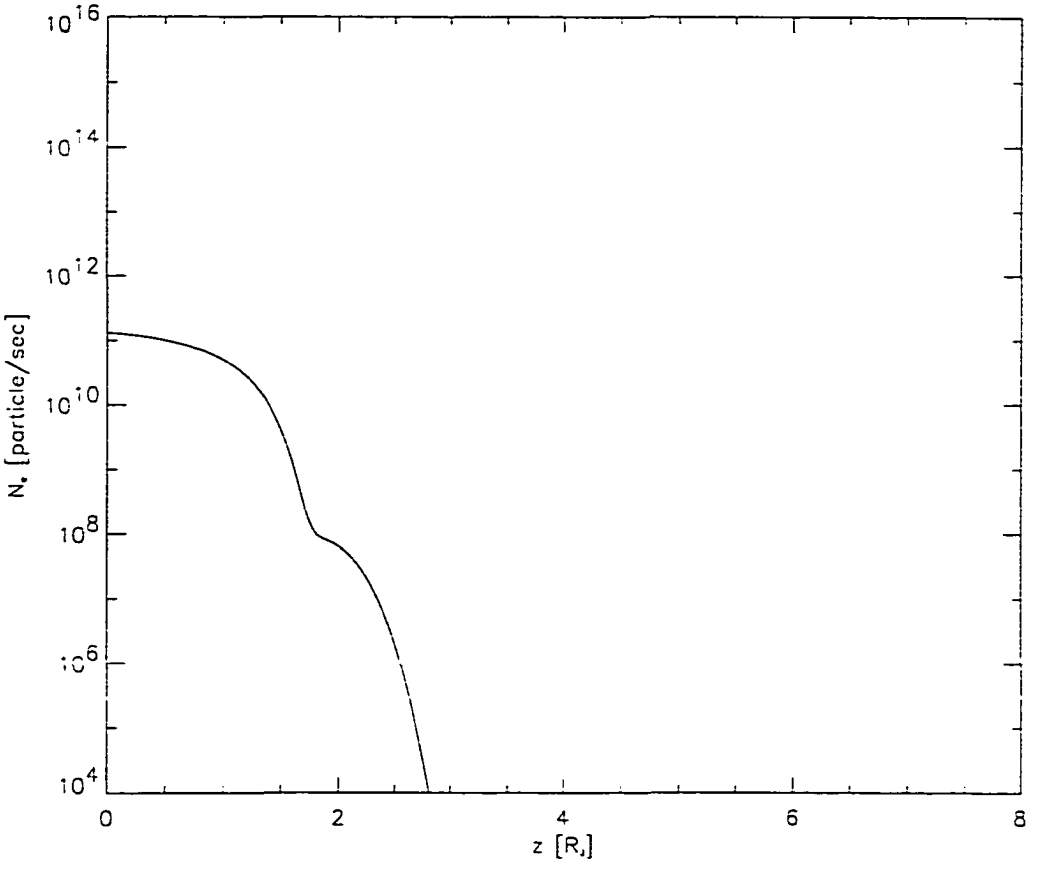


Figure 2.7: Rate of particle trapping by Fermi acceleration, per meter of distance along the Io flux tube, as a function of distance from Io.

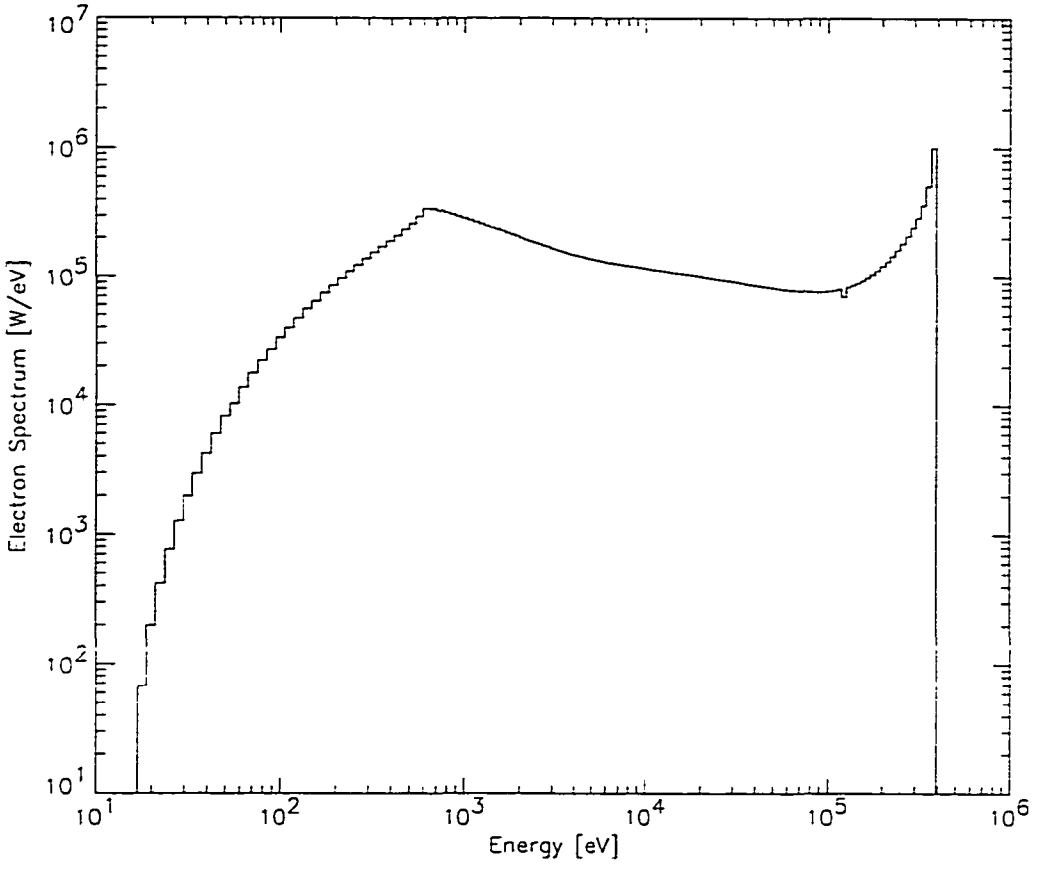


Figure 2.8: Energy spectrum of accelerated electrons

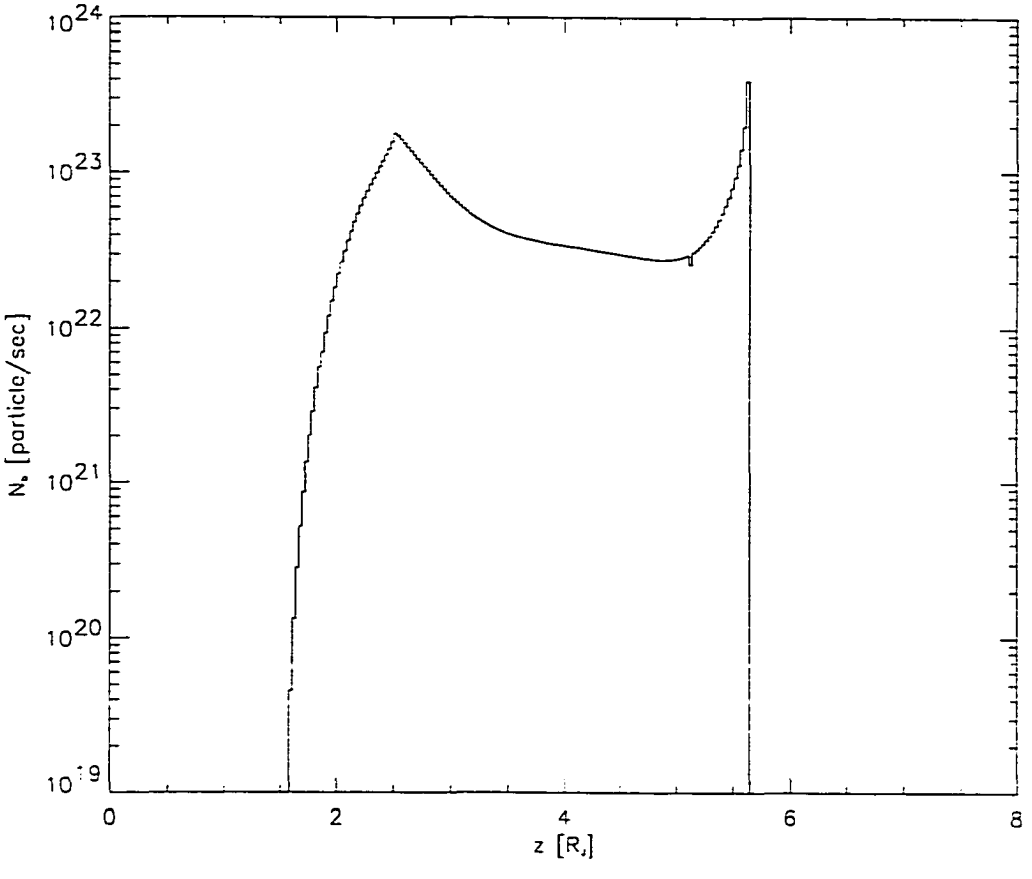


Figure 2.9: Rate of accelerated electron decoupling from the Alfvén wave, per meter of distance along the Io flux tube, as a function of distance from Io.

the non-uniform medium is the dominant process. Farther along the field line, the energy lost to accelerating the electron beam becomes suddenly becomes significant. Between 5.0 and 5.7 R_J ($\sim 42° - 47°$ latitude), the wave looses all of its remaining power to the electron beam. This accounts for the large number of particles which decouple near 5.7 R_J and the sharper, $\sim$ 400 keV peak in precipitating electron energies.

Of these electrons, virtually all are in the loss cone, assuming perpendicular temperatures are unchanged from the equatorial values of *Bagenal* [1994], a 4.9 eV core component and a 150 eV "hot" component. A small fraction of the accelerated electrons are outside the loss cone, in a distribution which is unstable and would be subject to the cyclotron maser instability. These particles account for less than 1% of the total power of the electron beam. They have energies between 0.7 and 20 keV, and a mean energy of 1.3 keV.

These results are uncertain, due to the approximations and assumptions required by the present model, but should be qualitatively correct. Using a numerically estimated Alfvén velocity, rather than the exponential model used in deriving the wave propagation, may introduce errors of order $\partial z_0/\partial z \sim 10\%$. The Alfvén velocity is not even approximately exponential near the equator. To some extent, this can be neglected, as noted above, by assuming WKB propagation until 1.5 R_J along the flux tube. However, the shape and spectrum of the wave packet may be affected. Since the present model largely focuses on the amplitude, rather than the shape, of the disturbance, this should not be a dominant source of error.

In solving equation 2.30, the $\partial b/\partial t$ term was neglected, based on an assumed form of the Io interaction. This may be incorrect, and the $\partial b/\partial t$ term may drive a significant portion of the wave packet. It can be shown that this term produces a more easily reflected disturbance, proportional to $V_A^{-3/2}$ rather than $V_A^{-1/2}$. As such, a significant $\partial b/\partial t$ contribution would result in a lower energy electron beam and a softer spectrum. However, unless the $\partial b/\partial t$ term is dominant (contrary to the models of Linker and Wolf-Gladrow), this reduction in power would be less than an order of magnitude.

The model is sensitive to the amplitude and power of the initial disturbance, and plasma densities at mid and high latitudes. These parameters are not well constrained by observations, although reasonable constraints may be placed on them. The effect of these uncertainties was tested by recalculating the electron beam power, for a variety of

parameters. The initial powers ranged from $5x10^{11}$ to $5x10^{12}$ W, the initial current density from 10^{-7} to $5x10^{-6}$ A/m^2. Modeled plasma densities covered the most extreme values consistent with the Voyager whistler wave observations [Crary et al., 1996]. This range also covers the uncertainties in extrapolating Voyager PLS observations to high latitudes [Bagenal, 1994]. Over this wide range of possible parameters, the predicted electron beam's power was between 10^{10} and $2.1x10^{11}$ W, with most parameters giving a power in the 0.5 to $1.5x10^{11}$ W range. The flux of precipitating particles ranged from $1.3x10^{24}$ to $2.4x10^{25}$. The greatest uncertainty is in the power and energy of electrons outside the loss cone. Their average energies ranged from 2.9 to 18 keV, and they represent between $1.2x10^{8}$ and $5.5x10^{9}$ W of power.

Finally, the treatment of the wave-particle interaction in this study is limited. In this model only imposes conservation of energy: The wave amplitude is decreased as electrons are accelerated, so that the total energy of the wave and electrons remains constant. In reality, the beam's own electric and magnetic fields would also affect the wave.

2.7 Comparison to Observations

Since the Alfvén wave itself is reflected and never reaches the high latitudes, this electron beam is the only Alfvén -related connection between and the Io-related, high latitude phenomena. The beam could directly produce an auroral spot simply by precipitating into the upper atmosphere. Measurements of the Io-related spot have been made in the far ultraviolet and at the 3.40 μm line of H_3^+. At 3.40 μm, the spot emitted "0.3 to 1.0% of the (variable) total flux from the aurora" [Connerney et al., 1993], implying a power of order 10^{11} W. In the far ultraviolet, the auroral spot emissions were 10% to 20% that of the main auroral oval, for an estimated spot power of "nearly 10^{11} W" [Clarke et al., 1996] and $2 - 3x10^{11}$ W [Prangé et al., 1996]. This is consistent with the power predicted by the present model, within the uncertainties discussed above.

The arrival time for the electrons is a function of their energy and the point at which they decouple from the Alfvén wave. The 100 keV electrons would reach the ionosphere 245 seconds after Io first disturbed the field line, the 1 keV electrons after 255 sec., and so on. This would imply a lag of 1.85° to 2.0° between the auroral spot and the foot of the field

line currently connected to Io. This is not in agreement with the initial observations of the auroral spot, which suggest a lag of 15° to 20° [Connerney et al., 1993]. However, more recent observations of the H_3^+ emissions show that the lag can not be accurately estimated without further refinements of the magnetic field model [Connerney et al., 1995]. Similarly, ultraviolet observation of the spot show either a 0° to 10° lag, with an estimated 5° to 10° uncertainty from the magnetic field model [Clarke et al., 1996], or a lag of less than 12° for all images of the aurora and under 4° in one case [Prangé et al., 1996].

In addition to the auroral spot, the electron beam would also be a power source for the decametric radiation. The exact mechanism producing the decametric emissions is unknown. Several possibilities have been suggested, but the cyclotron maser instability is the mechanism most frequently discussed in the literature. Recent observations of the rapid, "S burts" emissions are consistent with a cyclotron maser instability from 5.3 ± 2.2 keV electrons outside of the loss cone and traveling away from Jupiter's ionosphere. The instantaneous bandwidth and the duration of the bursts implies an instantaneous source region of under 20 km along the field lines [Zarka et al., 1996]. A similar but independent analysis of S bursts [Carr et al., 1997] agrees, although the electron velocity was found to be $1.64 \cdot 10^7$ m/s (765 eV).

The electron beam modeled in this chapter is the only part of the initial disturbance which reaches the high latitudes. The result of the disturbance is a pulse of electrons accelerated at the front of the Alfvénic disturbance. Those particles which mirror would be traveling away from Jupiter's ionosphere with roughly the correct loss cone distribution and energy range, 3-18 keV, and the bursty character need to account for the S-burst emissions. However, the power in the reflected particles may be insufficient. The total power radiated by Jupiter's decametric emissions is a function of the emission cone angle and width, variables that are not well-constrained by observations. Isotropic emission, which is certainly not accurate, sets an upper limit of $3x10^{11}$ W. A more realistic geometry, an emission cone with a 5° thickness, would require $6x10^9$ W [Carr et al., 1983]. The S-bursts represent roughly 10% of the decametric emissions, a radiated power of order $5x10^8$ W. But the efficiency of the cyclotron maser instability, even at saturation, is only a few percent. Therefore, electron beams with a power of over 10^{10} W are needed to explain the S-bursts. My $1.2x10^8$ to $5.5x10^9$ W power estimate is insufficient. However, that estimate is very sensitive to the electron

distribution. The model contains no perpendicular acceleration, and the electrons decouple from the wave with a velocity of $\sim 1x10^7$ m/s, requiring perpendicular velocities of a few times 10^6 m/s, or several times the thermal velocity, to be outside the loss cone. As a result, the reflected electrons are from the wings rather than the core of the electron distribution function. A two-component Maxwellian distribution was used in the above calculations, and this may significantly underestimate the number of the reflected electrons.

An alternative is to use a κ, or generalized Lorentzian, function rather than a Maxwellian for the electron distribution [Vasyliūnas, 1968]. Observations of the plasma waves by the Ulysses spacecraft have been analyzed to determine electron densities and temperatures in the Io torus. An anti-correlation between temperature and density was observed, and interpreted as evidence of a κ function distribution [Meyer-Vernet et al., 1995]. Although these measurements were made near L=7.5 to 8, rather than Io's L $\sim$ 6, I have used the Ulysses results, $T_e = 5.25eV$ and $\kappa = 2.4$, to illustrate the effect of a nonthermal tail. When the model calculations are repeated, using this κ distribution, the total power of the electron beam does not change significantly. The flux of electrons increases, typically by a factor of a few, since more electrons satisfy equation 2.53 when $V_A \gg v_{th}$. The location at which the electrons decouple from the wave shifts to lower latitudes, resulting in lower energy electrons. These two effects cause the total power of the beam to remain in the 0.5 to $1.5x10^{11}$ W range for most combinations of parameters. The effect on the reflected electrons is more extreme. Their average energy increases from 1 to 1.5 keV to between 4.5 and 7.5 keV. The power going into trapped electrons increased from under $6x10^9$ W to $2x10^9 - 2x10^{10}$ W. At the higher end of this predicted range, the trapped electron beams would have sufficient power to generate the S-bursts, assuming a conversion efficiency of roughly 2.5%.

2.8 Summary

In this chapter, I have investigated the initial Alfvénic disturbance caused by Io. I derived a model of the propagation of an Alfvénic wave packet as it travels through a non-uniform medium. My analysis shows that very little wave energy could travel from Io to Jupiter's ionosphere. Instead, most of the wave energy would be trapped in the Io plasma

torus.

I then considered the role of finite electron inertia, which results in non-zero, parallel electric fields. This effect is strong, since the perpendicular scale of the Io-genic waves is large compared to the plasma skin depth. Despite this, few to tens of volt potentials occur at the wave front, as the Alfvénic disturbance propagates into the low density region above the Io plasma torus. Electrons which encounter the wave experience Fermi acceleration. Although the field-aligned potential is tens of volts or less, the electrons are accelerated to high energies due to the increasing wave velocity. The wave velocity increases as the wave travels away from Io, while the velocity of an electron remains constant after encountering the wave. As a result, the electron will re-encounter the wave, and be accelerated many times. The important requirements for this acceleration mechanism are an Alfvén wave which propagates into a lower density plasma and which has an increasing group velocity.

I use this model to calculate the power and energy of the resulting energetic electrons. The results depend on the assumed initial wave power and current density and on density of the mid- and high-latitudes of the Io flux tube. I find that roughly 10% of the initial wave power, or 0.5 to 1.5×10^{11} W, is converted into energetic electron beams. These electrons have a wide range of energies, from a few hundred to a few hundred thousand eV. Almost all of these electrons would precipitate, resulting in the auroral spot at the foot of Io's flux tube. My calculations are consistent with the observed power of the spot.

A small fraction of the electrons would mirror. The flux of mirroring particles depends strongly on their perpendicular energy distribution. If Maxwellian electrons were accelerated, the mirroring electrons would represent a power of under 6×10^{9} W; if the electrons were in a κ distribution, this could be as high as 2×10^{10} W. These electrons would have a loss cone distribution and generate radio emissions through the cyclotron maser instability. The power I predict for electrons in a κ distribution is in reasonable agreement with the power radiated by decametric S-bursts and the efficiency of the this instability. In addition, my model predicts electron energies of 1 - 10 keV, in good agreement with observations of S-bursts, and the electrons would in a spatially confined pulse, consistent with the morphology of S-bursts.

I conclude that Io's auroral spot and the S-burst decametric emission are related. They are both the result of electrons accelerated by non-MHD effects at the front of an

Io-genic Alfvén wave packet. This acceleration process differs from that which produces Io's auroral arc and the bulk of the decametric emissions.

Chapter 3

Particle acceleration by a steady current system

3.1 Introduction

As I discussed in chapter 2, the extent of the decametric source region and the Io-associated UV aurora require continued particle acceleration at significant distances downstream from Io. The Galileo observations of a near-stagnant wake provide a likely source of field-aligned currents: The slowed flux tubes would be accelerated back up to the corotation velocity by a current loop connecting the wake and the jovian ionosphere [Pontius, 1995]. This requires a total current of order $5 \cdot 10^6$ Amps, and dissipate roughly 10^{12} W. However, the process of re-accelerating a slowed flux tube is generally assumed to be a MHD process [Hill and Pontius, 1998; Pontius, 1995]. While the re-acceleration of Io's wake can supply strong Birkeland currents, sufficient power and easy coupling to the high latitudes (unlike Io's initial Alfvén wave), the process by which particles are accelerated is unclear.

In this chapter, I will discuss two possible sources of acceleration due to quasi-static Birkeland currents. First, I will estimate the field-aligned potential drop in Jupiter's collisional ionosphere. Since this is the only region where the plasma has a finite, parallel conductivity, it is a natural place to look for field-aligned potential drops, and therefore particle acceleration. However, I will show that this process cannot result in significant particle acceleration. Application of the same model does suggest that acceleration in a collisional ionosphere is significant for other planets, especially Venus and Mars which have dense but weakly magnetized ionospheres.

In the absence of a MHD acceleration process, I examine the role of electrostatic double layers. These are kinetic structures which contain parallel electric fields, localized to a region of order a few Debye lengths. I will describe the basic character of double layers, illustrate the formation of multiple double layers along a given field line, and apply the results to the Io interaction. I find that multiple double layers are likely to form, that they will greatly reduce the current density in Io's wake (lengthening it to a size consistent with UV auroral observations) and would result in particle acceleration consistent with the required source of the bulk of the decametric emissions and with the bi-directional electron beams observed in Io's wake by Galileo.

3.2 Current closure within an ionosphere

In the collisional ionosphere, the DC parallel conductivity is finite. As a result, this is a plausible location for field-aligned potential drops. In this section, I will derive an analytic potential structure for an ionosphere, using an approximate model of the ionosphere and neutral atmosphere. I find that field-aligned potential drops are negligible unless the electron collision frequency is comparable to the electron cyclotron frequency at the base of the ionosphere. Since this is not the case for Jupiter, this process can be neglected. However, it may have applications to particle acceleration in weakly magnetized ionospheres such as those of Mars and Venus.

The potential structure in the ionosphere may be determined by requiring that $\vec{\nabla} \cdot \vec{J} = 0$. If the Hall conductivity is neglected, and the ionosphere is horizontally uniform,

$$\begin{aligned} 0 &= \vec{\nabla} \cdot \vec{J} \\ &= \vec{\nabla} \cdot (\sigma_{\parallel} E_z \hat{z} + \sigma_P E_x \hat{x}) \\ &= \sigma_{\parallel} \frac{\partial E_z}{\partial z} + E_z \frac{\partial \sigma_{\parallel}}{\partial z} + \sigma_P \frac{\partial E_x}{\partial x} \end{aligned} \tag{3.1}$$

Since $\frac{\partial}{\partial t} B = 0$ in a steady state, the electric field may be expressed as $-\vec{\nabla}\Phi$, although $\vec{\nabla}^2\Phi$ may not be zero. Since the above equations are linear, the imposed potential at the top of the ionosphere may be Fourier transformed, resulting in the equation,

$$0 = \frac{\partial^2 \Phi}{\partial z^2} + \frac{1}{\sigma_{\parallel}} \frac{\partial \sigma_{\parallel}}{\partial z} \frac{\partial \Phi}{\partial z} - k^2 \frac{\sigma_P}{\sigma_{\parallel}} \Phi \tag{3.2}$$

Current closure and the potential structure of interest occur near the altitude of peak ionospheric conductivity, roughly 1000 to 2000 km above the one bar level. There, the neutral density is less than 10^{-10} kg/m^3 [Seiff et al., 1996], or $3{\cdot}10^{10}$ cm^{-3} assuming the neutrals are primarily molecular hydrogen, and the neutral scale height is 200 km while the electron scale height is between 100 and 1000 km [Hinson et al., 1997], suggesting a plasma temperature of order 0.1 eV. The peak electron density is a few times 10^7 cm^{-3}. Observations show that ionospheric conditions on Jupiter are extremely variable, and the available data was obtained in the low and mid-latitude regions rather than the auroral ionosphere. I will use these data as an approximate starting point, but the final analysis will

retain these variables as free parameters.

For the expected neutral densities and plasma temperatures, assuming a collision cross section of $5{\cdot}10^{-19}$ m^2 and using $\nu_{\alpha,n} = Av_{th,\alpha}n_n$, the ion-neutral collision frequency would be under 25 s^{-1} and the electron-neutral collision frequency under 1000 s^{-1}. Since the magnetic field strength in the auroral ionosphere is over 600 μT, and the ion and electron gyrofrequencies over $6{\cdot}10^4$ and 10^8 Hz, respectively, $\nu_\alpha \ll \Omega_\alpha$. In this limit, the ionospheric conductivity is

$$\sigma_\parallel = \epsilon_0 \frac{\omega_{pe}^2}{\nu_{en}} \tag{3.3}$$

$$\sigma_P = \epsilon_0 \frac{m_e \omega_{pe}^2 \nu_{in}}{M_i \Omega_i^2} \tag{3.4}$$

The ratio of conductivities in equation 3.2 becomes

$$\begin{aligned} \frac{\sigma_P}{\sigma_\parallel} &= \frac{M_i \nu_{in} \nu_{en}}{m_e \Omega_{ce}^2} \\ &= \sqrt{\frac{M_i}{m_e}} \left(\frac{\nu_{ce}}{\Omega_{ce}} \right)^2 \end{aligned} \tag{3.5}$$

which is of order 10^{-9}.

To find an analytic solution to equation 3.2, I will assume that the neutral density is

$$n_n = n_{n,0} \exp\left(-\frac{z}{H}\right) \tag{3.6}$$

and the electron density is

$$n_e = n_{e,0} \exp\left(1 - e^{-z/H}\right) \tag{3.7}$$

This electron density differs from that of a Chapman profile,

$$n_e = n_{e,0} \exp\left((1 - \frac{z}{H} - \frac{1}{2} e^{z/H}\right) \tag{3.8}$$

in two respects. First, the increase in electron density is more gradual, in the region $z < 0$. Second, the electron density does not peak at $z = 0$. Instead, it continues to rise and approaches an asymptotic value as $z \to \infty$. This electron density profile was chosen to produce an analytic solution to equation 3.2. However, the results are physically reasonable. The observed ionosphere of Jupiter is broader and extends to higher altitudes than predicted

by equilibrium photoionization and recombination [Strobel and Atreya, 1983]. This may be a result of vertical transport and/or ionization by precipitating particles. The increased vertical extent may be greater in the auroral latitudes, due to magnetospheric heating. As a result, equation 3.8 is reasonably accurate for $z < 5H$. At higher altitudes, this electron profile is not accurate but does not affect the resulting solution to equation 3.2. As discussed below, current closure occurs below $z = 5H$. Above this altitude, the Pedersen conductivity is essentially zero and the parallel current and the potential have nearly reached their asymptotic value.

From this electron and neutral density profile, it follows that

$$\frac{\sigma_P}{\sigma_\parallel} = \frac{\sigma_{P,0}}{\sigma_{\parallel,0}} e^{-2z/H} \tag{3.9}$$

$$\frac{1}{\sigma_\parallel}\frac{\partial}{\partial z}\sigma_\parallel = \left(1 + e^{-z/H}\right) H^{-1} \tag{3.10}$$

where the subscript, 0, indicates the conductivity at $z = 0$. Inserting these relations into equation 3.2 and making the change of variables, $w \equiv e^{-z/H}$ results in

$$\frac{\partial^2}{\partial w^2}\Phi - \frac{\partial}{\partial w}\Phi - \eta^2\Phi = 0 \tag{3.11}$$

where $\eta \equiv kH\sqrt{\sigma_{P,0}/\sigma_{\parallel,0}}$. The solution to this equation is

$$\Phi = \Phi_0 \exp\left[\frac{1}{2}\left(1 - \sqrt{1+4\eta^2}\right) e^{-z/H}\right] \tag{3.12}$$

which goes to zero as $z \to -\infty$ and to Φ_0 as $z \to \infty$. The second solution is not of interest, since it goes to infinity as $z \to -\infty$.

As can be seen from equation 3.12, field-aligned potential drops do occur, but they occur at a depth of approximately $ln\left[\left(\sqrt{1+4\eta^2} - 1\right)/2\right] H$. Assuming that 1/k is roughly width of the Io flux tube, $\approx$ 200 km when mapped to Jupiter's ionosphere, and the neutral scale height along auroral field lines is similar to its equatorial value, $\eta \approx 5 \cdot 10^{-5}$. As a result, the potential drop occurs near $z = -18H$. This is well below the ionosphere and deep in the collisional atmosphere. Therefore, this potential drop is irrelevant to particle acceleration.

If η of order 0.1 or greater, the potential drop could occur in the ionosphere and within a few mean free paths of the exobase. A small fraction of the ionospheric electrons could escape the atmosphere after having been accelerated through the potential drop. Although this does not apply to Jupiter, it could be significant for planets with collisional ionospheres but weak (or nonexistent) magnetic fields, such as Venus and Mars.

The asymptotic limits of equation 3.12 and 3.8 determine the effective conductivity of the ionosphere. As $z \to \infty$,

$$\begin{aligned} J_z &= -\sigma_{\parallel} \frac{\partial}{\partial z} \Phi \\ &= \Phi_0 k^2 H \sigma_{P,0} e \left(\frac{1-\sqrt{1+4\eta^2}}{2\eta^2} \right) \\ &\to -\Phi_0 k^2 H e \sigma_{P,0} \text{ as } \eta \to 0 \end{aligned} \tag{3.13}$$

Integrating over x gives the current per unit length,

$$\begin{aligned} j &= \Phi_0 k H \sigma_{P,0} e \left(\frac{1-\sqrt{1+4\eta^2}}{2\eta^2} \right) \\ &\equiv \Sigma_{ionosphere} \Phi_0 k \end{aligned} \tag{3.14}$$

$$\Sigma_{ionosphere} \equiv H \sigma_{P,0} e \left(\frac{1-\sqrt{1-4\eta^2}}{2\eta^2} \right) \tag{3.15}$$

In theoretical models of magnetosphere-ionosphere coupling [Hill, 1979; Huang and Hill, 1989; Pontius, 1997] the height integrated Pedersen conductivity is generally used. For the electron and neutral densities assumed here, the height integrated Pedersen conductivity is,

$$\Sigma_P \equiv \int_{-\infty}^{\infty} dz \sigma_P = eH\sigma_{P,0} \tag{3.16}$$

which differs from the ionospheric conductance I have derived by a factor of $(\sqrt{1+4\eta^2} - 1)/2\eta^2$. Since $\eta \approx 5 \cdot 10^{-5}$, the two conductances differ by less than 1 part in 10^{10}. The differences only become significant as η approaches unity. Again, this is not significant for Jupiter, but may be important to ionosphere-solar wind coupling in the magnetospheres of Mars and Venus.

3.3 Double Layers

Since steady electron acceleration does not occur in the ionosphere, I will discuss the properties of double layers and their application to the Io interaction. In this section, I summarize the properties of double layers, suggest a somewhat different approach to analytically modeling their structure. Based on this analysis, I conclude that very strong magnetospheric double layers are unlikely, but that multiple double layers of moderate amplitude would form along field lines connected to Io and Io's wake.

Double layers are electric potential drops along a magnetic field line and are typically a few to tens of Debye lengths in size. Intuitively, the ability of the plasma to sustain a potential drop can be thought of as dynamical, or anti-, charge shielding [Hansen et al., 1996]. Consider a beam of electrons passing through a region of positive electric potential. The electrons are accelerated as they enter the region and decelerated as they leave it. In a steady state, the electron flux, $n_e < v >$, must be constant. So in a region of positive potential, the electron density must decrease. Similarly, the ions entering the region would be slowed and their density would increase. Thus, in this region, the plasma has a positive charge density. This may either sustain/cause the potential structure (as in the case of a double layer) or it may cause an instability (e.g. the Buneman instability, [Galeev et al., 1981]).

Double layers are described as "strong" or "weak", depending on the potential across them. A double layer is "strong" if $e\Phi_0 > kT_e$. Extremely strong double layers, $e\Phi_0 \gg kT_e$, have been proposed as an acceleration mechanism in solar flares and astrophysical objects [Raadu, 1989; Volwerk, 1993a]. While double layers of order or stronger than $e\Phi_0 \sim 25kT_e$ are easily created in laboratory experiments, very strong double layers have not been observed in the magnetosphere [Borovsky, 1993]. Double layers have been seen by the Viking [Bostrom et al., 1988], Freja and Polar spacecraft, but they are not especially strong, with potentials of under 5 V [Mälkki et al., 1993], in a region where the background electron temperature may be of order a few volts. Even in laboratory experiments, multiple, lower amplitude double layers have been observed to form rather than one, very strong double layer. For example, Guyot and Hollenstein, 1983, reported experiments that generated multiple double layers with $e\Phi_0 = 10 - 12kT_e$ each, when a 50 V potential was applied

across a plasma. In experiments, the separation between multiple double layers is found to be of order a few thousand Debye lengths [Guyot and Hollenstein, 1983; Raadu, 1989].

The conditions under which a double layer will form are described by the Bohm criteria, or the equivalent threshold for the Buneman instability,

$$< v_e >= \left(\frac{J_{\|}}{en_e}\right)^2 > \frac{2k}{m_e}(T_e + T_i) \tag{3.17}$$

Galeev et al., 1981, showed that the nonlinear stage of the Buneman instability results in a density cavity. If an external process is driving a field-aligned current and/or a potential difference across the plasma, this density cavity would evolve into a double layer [Galeev et al., 1981; Raadu, 1989].

In addition to accelerating particles, double layers can be a significant source of plasma waves. The accelerated particles are frequently subject to various beam instabilities. In the laboratory, high frequency waves are seen often measured within a few hundred to a thousand Debye lengths the high potential end of double layers [Guyot and Hollenstein, 1983; Volwerk, 1993b], as well as low frequency waves on the opposite side of the double layer [Raadu, 1989]. In the Earth's magnetosphere, Viking frequently observed electrostatic ion cyclotron waves in the vicinity of weak double layers, but it is unclear if this is a causal relationship [Mälkki et al., 1993].

The details of a double layer's structure can be calculated using Vlasov's and Poisson's equations [Raadu, 1989; Schamel et al., 1983]. In the following illustration, I will largely follow the technique described by Schamel and Bujarbarua, 1983, but apply significantly different assumptions about the plasma. In this approach, the distribution function of particles moving towards the double layer is specified; the ion and electron densities are determined as a function of electric potential, then $n_i(\Phi) - n_e(\Phi)$ is used to solve Poisson's equation.

Assuming a steady state and neglecting motion perpendicular to the magnetic field, Vlasov's equation is solved by

$$\begin{aligned} f_\alpha(v,\Phi) &= f_\alpha\left(\sqrt{v^2 - \tfrac{2q\Phi}{m}}, \Phi = 0\right) \\ &= f_\alpha\left(\sqrt{v^2 - \tfrac{2q(\Phi-\Phi_0}{m}}, \Phi = \Phi_0\right) \end{aligned} \tag{3.18}$$

where f is the distribution of a given component, as a function of velocity and electric potential. The electric potential is defined to be zero on one side of the double layer and is Φ_0 on the opposite side.

The electron distribution may be divided into two components: Particles approaching the double layer from the $\Phi = 0$ side, with the distribution function f_{e-}, and those approaching from the $\Phi = \Phi_0$ side, f_{e+}. The electrons approaching from the $\Phi = 0$ side will be accelerated by the double layer's electric field. Those approaching from the other side will be slowed and if they lack sufficient energy, reflected. By specifying $f_{e-}(\Phi = 0)$ and $f_{e+}(\Phi = \Phi_0)$, the electron density as a function of Φ can be determined. The ion density can be similarly determined. Figure 3.1 illustrates these components.

In their analytic models of double layers, Schamel and Bujarbarua, 1983, used a convected Maxwellian distributions for $f_{e-}(\Phi = 0)$ and $f_{e+}(\Phi = \Phi_0)$. However, they treated the densities, temperatures and beam velocities as free parameters. In this case, there are no limits on the strength of a double layer (i.e. $e\Phi_0 >> T_e$ is possible.) These solutions required temperature or density contrasts across the double layer. I do not consider this a reasonable approach to magnetospheric double layers: These distributions are an externally imposed boundary condition, not free parameters. In addition, at the time the double layer forms, the plasma above and below the double layer has the same density and temperature. A density or temperature contrast across a double layer is only reasonable if the double layer is able to significantly alter its environment (e.g. by heating the plasma on one side.) This is quite possible in laboratory experiments, but less likely in a magnetospheric plasma.

The formation of double layers is related to the Buneman instability [Galeev et al., 1981; Raadu, 1989]. In a nonlinear analysis of the Buneman instability, Galeev et al., (1981) showed that it occurs on a time scale of $\pi\sqrt{m_i/m_e}L/(2 < v_+ >)$ where $< v_+ >$ is the average velocity of electrons entering the region and L is a characteristic length scale. For an instability which would produce a double layer, $< v_+ >$ would be slightly greater than the thermal velocity, and L of order the Debye length. It follows that a typical time for a double layer to form would be of order $\sqrt{m_i/m_e}\omega_{pe}^{-1}$. For the mid- to high latitudes of the Io flux tube, with $n_e \sim 10cm^{-3}$ and protons as the dominant ions, this is roughly $2 \cdot 10^{-4}$ sec. The time for a double layer to alter its environment might be of order the electron bounce time. In the absence of scattering, this is the time required for an electron to return to the

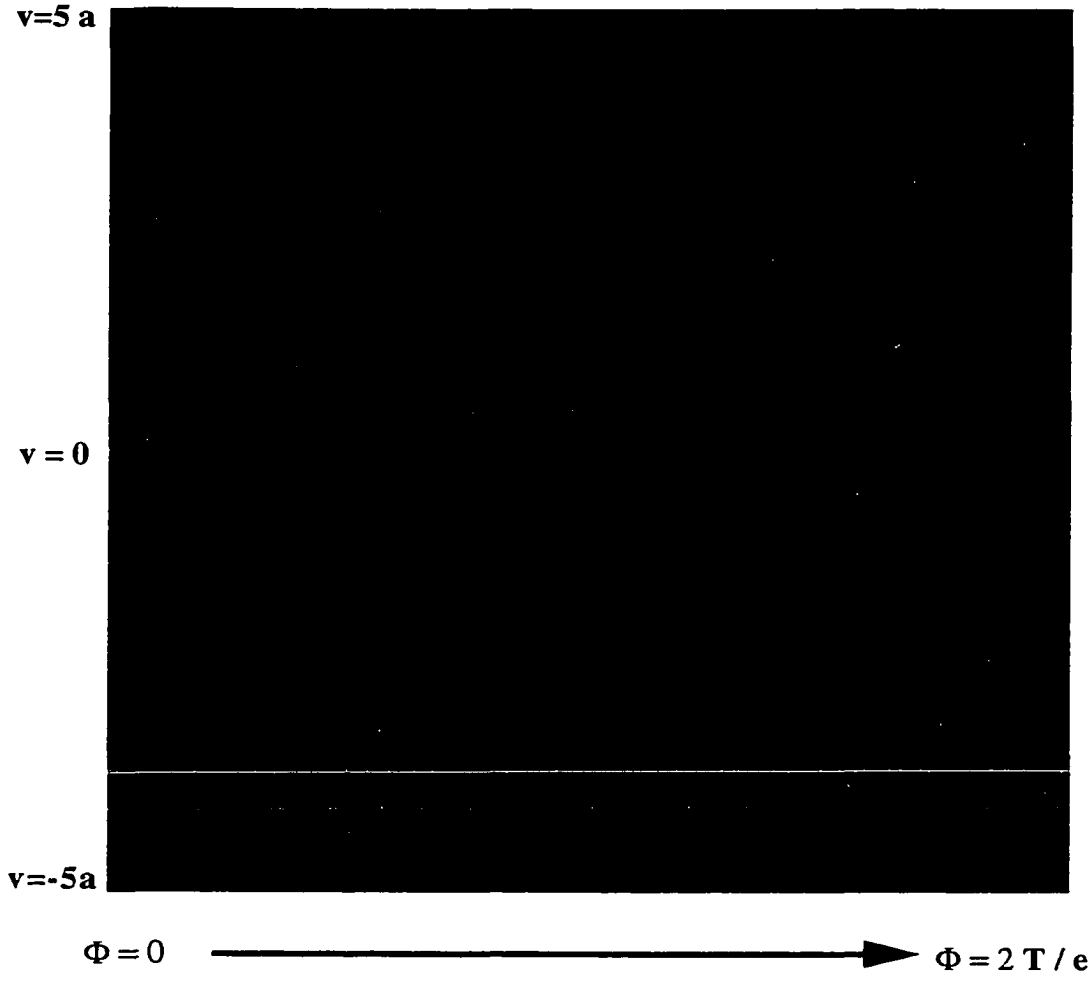


Figure 3.1: The components of the electron distribution in a double layer. The electrons entering from the low potential side, f_-, are shown in blue and are accelerated by the electric fields. Those entering from the high potential side, f_+, are shown in green and are slowed. Depending on their initial energy, they may either be reflected or stream through the double layer.

double layer, i.e. for the flux of particles out of the region to mirror and become a flux of particles into the region. The bounce time for thermal (5 eV) electrons along Io's flux tube is around 1300 seconds. Even 200 keV electrons have a bounce time, $\sim$ 6 seconds, which is much longer than the time required to form a double layer. Lysak et al., 1982, pointed out this difference between laboratory and magnetospheric double layers, but did not pursue its implications. Since double layers would form in a uniform plasma and would not have sufficient time to make their environment non-uniform, I assume that the temperature and density of the plasma must be the same on both sides of the double layer. As discussed below, when I apply this assumption to an analytic model of double layers [Schamel et al., 1983] I obtain significantly different results, and an upper limit on the strength of double layers.

I will assume that the plasma is composed of protons and electrons, which is probably the case for Jupiter's mid- and high latitudes, and that the ion and electron temperatures are equal. Ulysses observations near L=7.5 showed that electron temperatures increase with latitude [Meyer-Vernet et al., 1995]. So, although the ion and electron temperatures are not equal within the plasma torus, this assumption is more plausible in the mid- and high latitudes. I will also assume that neither the ions or electrons have any beam velocity. To simplify the notation, I use non-dimensional variables, so that

$$\begin{aligned} n &\rightarrow n/n(\Phi = 0) \\ \Phi &\rightarrow e\Phi/kT_e \\ v &\rightarrow v/\sqrt{2kT_e/m_e} \\ x &\rightarrow x/\lambda_D \end{aligned} \tag{3.19}$$

The electron distribution functions are

$$\begin{aligned} f_{e+}(v, \Phi = \Phi_0) &= \begin{cases} \frac{A}{\sqrt{\pi}} e^{-v^2} & v < \sqrt{\Phi_0} \\ 0 & v > \sqrt{\Phi_0} \end{cases} \\ f_{e-}(v, \Phi = 0) &= \begin{cases} \frac{B}{\sqrt{\pi}} e^{-v^2} & v > 0 \\ 0 & v < 0 \end{cases} \end{aligned} \tag{3.20}$$

where A and B are constants relating the densities of the two components. The distribution

function, f_{e+} extends to $v = \sqrt{\Phi_0}$ in order to include both the particles approaching the double layer from the $\Phi = \Phi_0$ side and those which have approached from this side and been reflected, see figure 3.1. Inserting equation 3.18 and integrating over velocity, the electron density is

$$n_e(\Phi) = \left\{ A\left[1 + erf\left(\sqrt{\Phi}\right)\right] e^{(\Phi - \Phi_0)} + Berfc\left(\sqrt{\Phi}\right) e^{\Phi} \right\} \tag{3.21}$$

A and B are determined by the requirement that $n_e(\Phi_0) = n_e(0) = 1$, and the proton density is calculated in the same manner. This gives the charge density as a function of potential. In non-dimensionalized form, Poisson's equation becomes

$$n_e(\Phi) - n_i(\Phi) = \frac{\partial^2}{\partial s^2}\Phi \tag{3.22}$$

which may be solved by introducing the Sagdeev potential,

$$\begin{aligned} V(\Phi) &\equiv -\tfrac{1}{2}\left(\tfrac{\partial \Phi}{\partial x}\right)^2 \\ \tfrac{\partial V}{\partial \Phi} &= \tfrac{\partial^2 \Phi}{\partial s^2} = n_e(\Phi) - n_i(\Phi) \end{aligned} \tag{3.23}$$

The charge density can then be numerically integrated to give $V(\Phi)$ potential and then

$$s(\Phi) = \int_0^{\Phi} \frac{d\Phi'}{\sqrt{-2V(\Phi')}} \tag{3.24}$$

Although the Sagdeev potential was introduced as a mathematical convenience, it can be related to the energy density of the electric field [Raadu, 1989].

To be an equilibrium solution, a double layer must satisfy several conditions. At $\Phi = 0$ and $\Phi = \Phi_0$, both the charge density and the electric field must be zero. This is necessary for the double layer to be an isolated structure within an electrically neutral plasma. The formulation of ion and electron densities, above, automatically assures these conditions. In addition, the Sagdeev potential must be negative for $0 > \Phi > \Phi_0$. This follows from equation 3.24, and the fact that s is a real number. However, there is a more intuitive interpretation of this condition. If $V(\Phi_1) \rightarrow 0$ at some potential, Φ_1, then $E = -\frac{\partial \Phi}{\partial s} \rightarrow 0$ as well. By definition of the Sagdeev potential, the plasma would also be charge neutral at $\Phi_1 = 0$. Such a solution would consist of two, isolated, electrostatic structures: One in which the potential went from zero to Φ_1 and a second in which the potential went from Φ_1

to Φ_0. Since $\frac{\partial \Phi}{\partial s} = \frac{\partial^2 \Phi}{\partial s^2} = 0$, the two structures could be separated by an arbitrary distance. That is, the solution would be two, independent double layers along the same field line.

Figure 3.2 shows a solution to the above equations, for the case $e\Phi_0 = 2kT_e$. Near the center of the double layer, there is a decrease in electron and ion densities. Since the density depressions are not co-located, this results in two regions of non-zero charge: An ion excess for $s < 0$ and and electron excess for $s > 0$. The electric potential smoothly and monotonically increases from 0 to Φ_0. The transition from one to two double layers is illustrated in figures 3.3 and 3.4. As Φ_0 increases from 2 to 6 kT_e/e, the shape of the double layer does not change in any fundamental manner. It becomes thicker, and the two layers of opposite charge move farther apart, but the general character does not change. At $\Phi_0 \sim 7kT_e/e$, two additional layers of charge appear, one weakly negative at $-3 < s < 0$ and a symmetric, positive layer. As Φ_0 increases above $7kT_e/e$, the amplitude of the secondary layers increases, and by $\Phi_0 = 10.95kT_e/e$, two nearly-isolated double layers have developed. The layers of charge have separated into two physically separated pairs and the potential changes sharply in two distinct regions, but $E = -\frac{\partial P}{\partial s} \equiv 0$ at $s = 0$. In addition, as $\Phi_0 \rightarrow 10.96kT_e/e$, $V(\Phi = \Phi_0/2) \rightarrow 0$, which is my criteria (discussed above) for the existence of multiple double layers.

The formation of multiple double layers along the same field lines has interesting implications for the Io interaction and for terrestrial observations. My analysis was not general enough to draw reliable conclusions, since I only considered only one distribution of approaching particles. Differing electron and ion temperatures, beam components, etc. should also be examined. Despite this, my analysis appears to indicate that multiple double layers are more likely than very strong double layers, assuming that the double layer is in an externally imposed, uniform environment. This may explain why strong double layers have not been seen in the Earth's magnetosphere. My analysis, magnetospheric observations [Mälkki et al., 1993] and laboratory experiments which produced multiple double layers [Guyot and Hollenstein, 1983] all involve double layers of moderate strength, $e\Phi_0 \sim kT_e$.

The possibility of multiple double layers also allows me to avoid many details of their dynamics in the jovian magnetosphere. I will assume that they form where ever the current density violates the Bohm criteria (equation 3.17). Multiple double layers might occupy a few thousand Debye lengths [Guyot and Hollenstein, 1983] and involve a potential

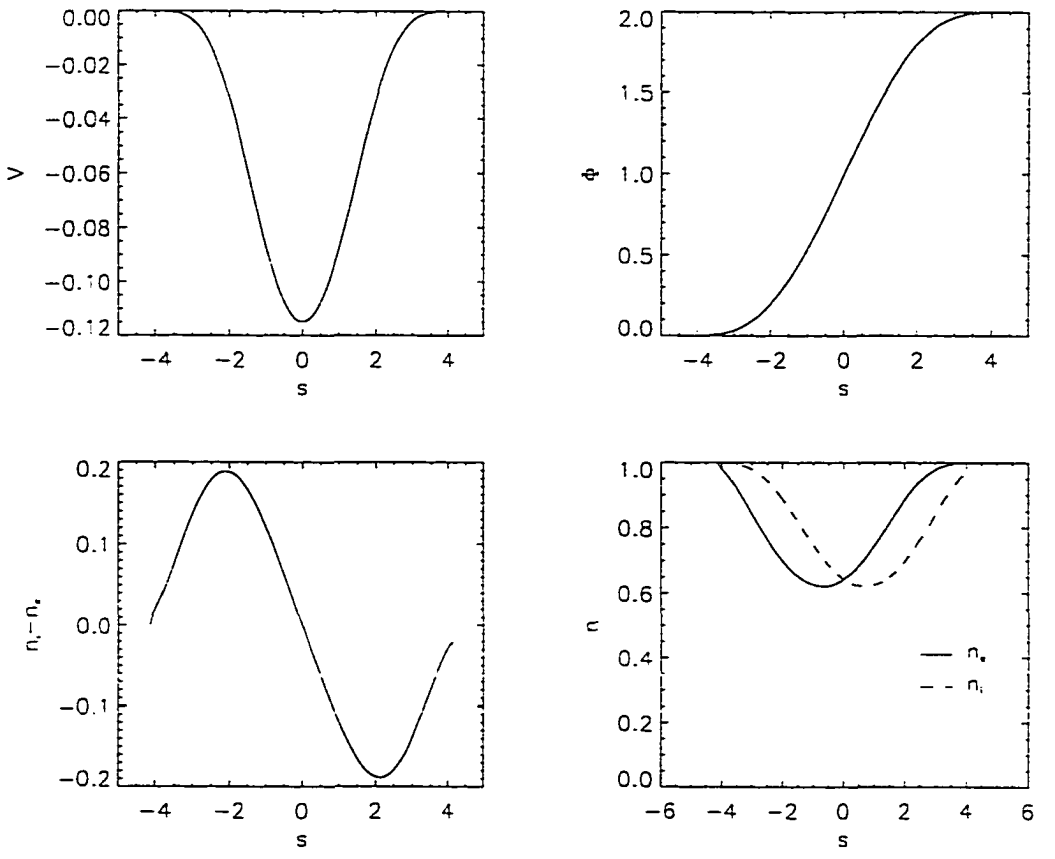


Figure 3.2: An example of a double layer solution. The double layer is constructed with a potential of $e\Phi_0 = 2kT_e$ across it. The Sagdeev potential, V, the real potential, Φ, the charge density and the ion and electron densities are shown as a function of position.

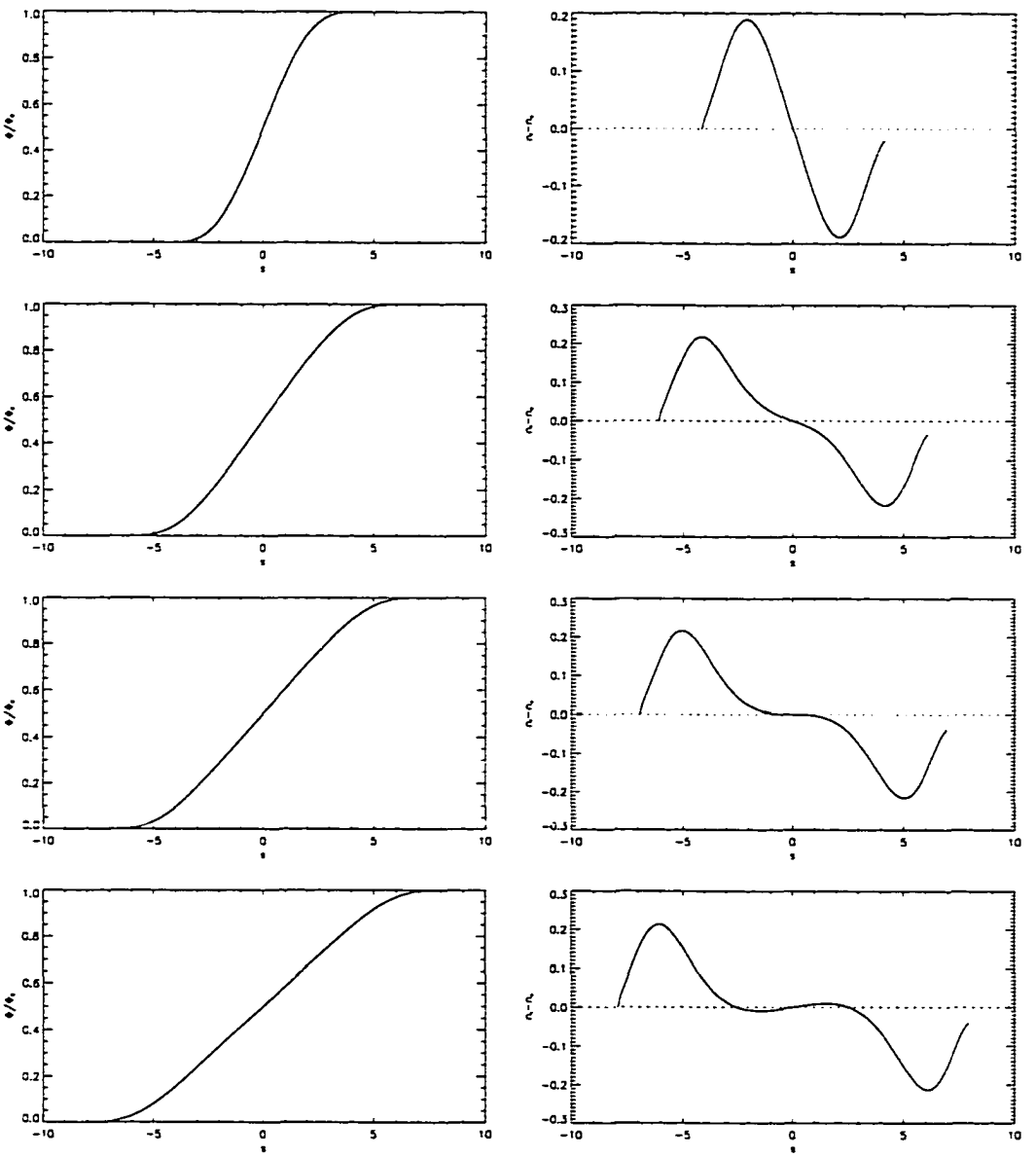


Figure 3.3: A series of double layers, showing the transition from a single double layer to multiple layer solutions. The Sagdeev potential and charge density are shown for eight different values of Φ_0. $e\Phi_0 = \{2, 5, 6, 7\}\, kT_e$.

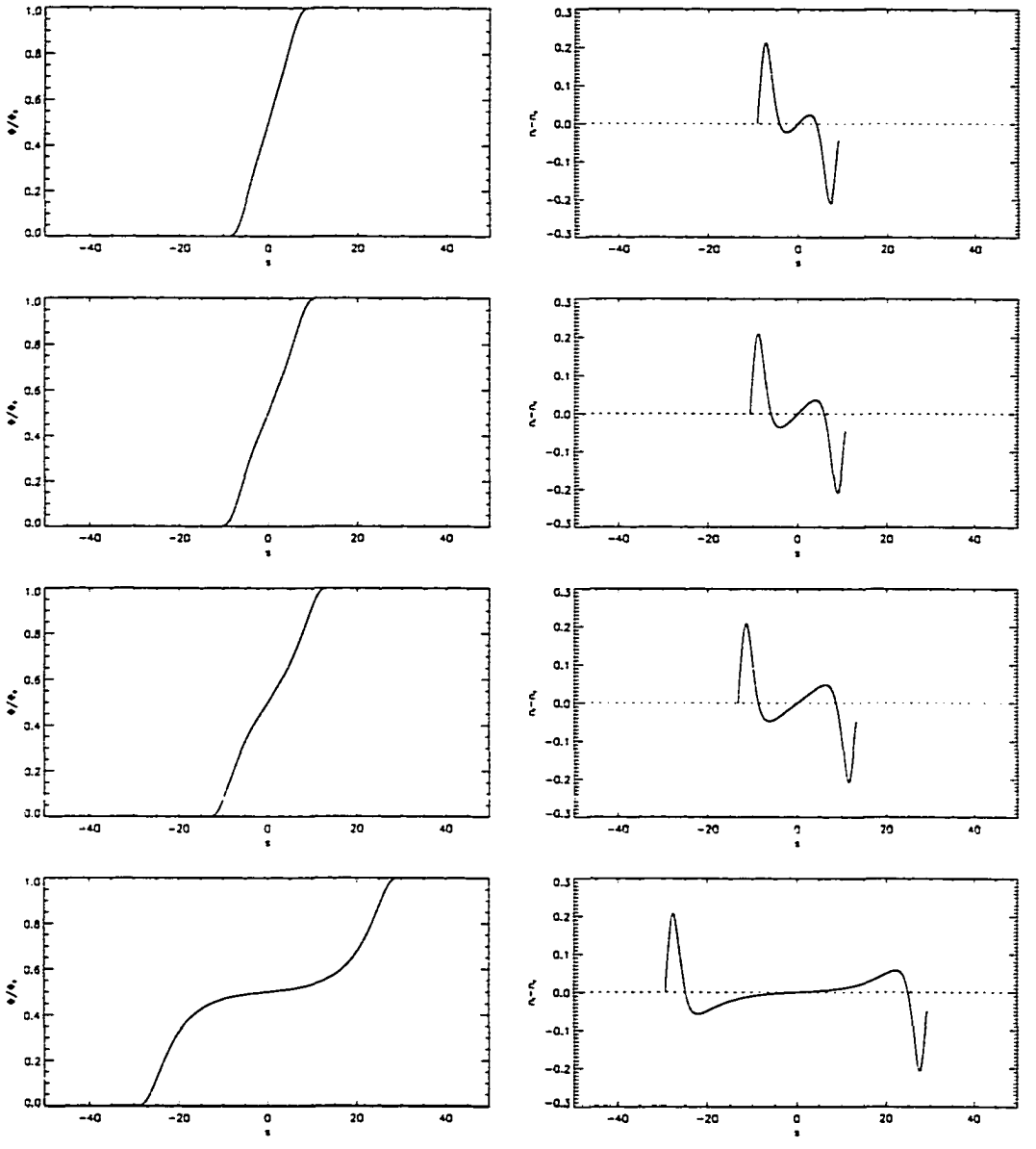


Figure 3.4: A continuation of figure 3.3 showing double layers with $e\Phi_0 = \{8, 9, 10, 10.95\}$ kT_e. Note the change in the horizontal scale from figure 3.3

drop of order the electron temperature. Since the potentials in question may be as high as 220 kV, and the electron temperature as low as 5 eV, this implies that thousands of double layers might form, occupying a region a few million Debye lengths in size. The potential across the double layers would reduce the potential applied across the ionosphere of Jupiter. This would reduce the current density along the Io flux tube. I will assume that, given sufficient room (i.e. a few million Debye lengths in which the Bohm criteria would otherwise be violated), enough double layers will form to reduce the current. In effect, that the double layers are a form of anomalous resistivity that limits the current density to $J_{||} = en\sqrt{2k\left(T_e + T_i\right)/m_e}$. It is worth noting that the double layers do not supply particle energy, and would not cause acceleration in the absence of an external source of energy and current. Their role is (1) to reduce the potential across the ionosphere and (2) to convert current carried by a large number of electrons at a low velocity, to an equal current carried by a small number of electrons at a high velocity.

3.4 Double Layers along the Io flux tubes

If the currents along the Io flux tube settled into a steady equilibrium before the field lines are advected past Io and no double layers form, the magnitude of the currents would depend only on the conductivities of Io and Jupiter. The height integrated current closing through Io is given by

$$j_{Io} = \Sigma_{Io}\Phi/2R_{Io} \tag{3.25}$$

where Φ is the potential across Io, Σ_{Io} the height integrated Pedersen conductivity and j the current, per unit length in the $\vec{E} \times \vec{B}$ (azimuthal) direction. Φ is determined by the corotation electric field, and the currents within Jupiter's ionosphere

$$\Phi = 2R_{Io}E_{cr} - 2R_{foot}j_{foot}/\Sigma_J \tag{3.26}$$

where the subscript, *foot*, refers to the current per unit length and radius of the flux tube, at its foot in the jovian ionosphere. In a symmetric, dipole field,

$$\begin{aligned} j_{foot} &= j_{Io} L^{3/2}/2 \\ R_{foot} &= R_{Io} L^{-3/2}/\sqrt{4-3/L} \end{aligned} \tag{3.27}$$

The factor of two in the equation for j_{foot} is a result of closure in the northern and southern hemispheres, i.e. half of j_{Io} closes in the north and half in the south. Combining these equations,

$$j_{Io} = E_{cr} \frac{2\sqrt{4-3/L}\Sigma_{Io}\Sigma_J}{\Sigma_{Io} + 2\sqrt{4-3/L}\Sigma_J} \tag{3.28}$$

This differs from the commonly stated form [Hill et al., 1983] by a factor of $\sqrt{4-3/L}$, which reflects the difference in the shape of the flux tube at the foot versus Io's location. If the current flowed in linear sheets, the total current would be $j_{Io}l$, where l is the length of the current system in the direction of flow, and the parallel current density would be j_{Io}/d where d is the thickness of the current sheets.

Numerically, $E_{cr} = 0.113V/m$ and $L = 5.9$. The other variables are less well known, but $\Sigma_J \sim 1\mho$ is consistent with observations of the equatorial ionosphere and $\Sigma_{Io} \gg \Sigma_J$ [Strobel and Atreya, 1983; Summers and Strobel, 1996]. The length of the current system would be of order R_{Io} and the thickness of the current sheets is related to the scale height of Io's ionosphere, perhaps 200 km, although variable, non-uniform and dependent on longitude. Inserting these values, I find that

$$\begin{aligned} j_{Io} &\sim 0.42A/m \\ J_{\parallel} &\sim 2 \cdot 10^{-6} A/m^2 \\ I &\sim 7.6 \cdot 10^5 A \end{aligned} \tag{3.29}$$

As noted earlier, high parallel current densities would drive instabilities and result in the formation of double layers in the high latitudes. These double layers will reduce the potential across the foot of Io's flux tube and prevent current densities $J_{\parallel,Io} > J_{crit}$, where J_{crit} threshold current for the Buneman instability. Since the currents are field-aligned, and the size of the Io flux tube changes with latitude, current density is a function of latitude. I

define J_{crit} in terms of the current density when mapped back to the equatorial plane,

$$J_{crit} = en_e\sqrt{\frac{2k}{m_e}(T_e + T_i)}\frac{B_0}{B} \tag{3.30}$$

Figure 3.5 shows this value as a function of distance along the Io flux tube. A current density under 10^{-8} A/m^2 would not result in any double layers. For a higher current density, double layers would form near the foot of the flux tube. However, the region in which $J > J_{crit}$ could be small and the number of double layers which could form would be limited. Multiple double layers in laboratory experiments are typically separated by a few thousand Debye lengths. If this separation also applies to the Io flux tube, then the number of double layers along the flux tube would depend on the the length of the field line where $J > J_{crit}$ divided by a few thousand Debye lengths. I assume the typical potential across a double layer would be a few to ten times the electron temperature, perhaps 50 eV each. To have a significant impact on the Io interaction, the total potential drop across the double layers would have to be of order the induced potential across Io, or 440 kV. This suggests that a few thousand double layers would be involved. With a separation of a few thousand Debye lengths, this implies that the field line would have to satisfy $J > J_{crit}$ for of order ten million Debye lengths, in order for the double layers to be significant. There exists a ten million Debye length long region in which J_{crit} is under $3 \cdot 10^{-8}$ A/m^2 or $1.5 \cdot 10^{-7}$ A/m^2, depending on the assumed temperatures. To illustrate the effects of the double layers, I assume the double layers either have no effect on the interaction (when $J < J_{max}$, for some J_{max}) or what ever potential is required to keep $J = J_{max}$. I adopt $J_{max} = 1 \cdot 10^{-7}$ A/m^2. Equations 3.25 and 3.26 then become

$$\begin{aligned} j_{Io} &= J_{max}d = \Sigma_{Io}\Phi/2R_{Io} \\ \Phi &= 2R_{Io}E_{cr} - \frac{dR_{Io}J_{max}}{\Sigma_J\sqrt{4-3/L}} - 2\Phi_{DL} \end{aligned} \tag{3.31}$$

where Φ_{DL} is the potential of the double layers along a given field line. As a result,

$$\Phi_{DL} = R_{Io}\left(E_{cr} - J_{max}d\frac{\Sigma_{Io} + 2\Sigma_J\sqrt{4-3/L}}{2\Sigma_{Io}\Sigma_J\sqrt{4-3/L}}\right) \tag{3.32}$$

Since the total current is dlJ_{max}, the power dissipated by the double layers is

$$P = dlJ_{max}\Phi_{DL} = dlR_{Io}\left(E_{cr}J_{max} - J_{max}^2 d\frac{\Sigma_{Io} + 2\Sigma_J\sqrt{4-3/L}}{2\Sigma_{Io}\Sigma_J\sqrt{4-3/L}}\right) \tag{3.33}$$

The dissipated power would be greatest if

$$J_{max} = \frac{\Sigma_{Io}\Sigma_J\sqrt{4-3/L}}{d(\Sigma_{Io} + 2\Sigma_J\sqrt{4-3/L})}E_{cr} \sim 1\cdot 10^{-6} A/m^2 \tag{3.34}$$

This power would go into accelerating electron and ion beams, and, indirectly, generating plasma waves through beam-related instabilities.

Using the values assumed above gives $\Phi_{DL} \sim 98$ kV. The total current would be very small, of order $3.7\cdot 10^4$ A, and the power dissipated in the double layers would be $7\cdot 10^9$ W. Compared to the magnetic perturbation observed by Voyager, $3.7\cdot 10^4$ A is an extremely small current. This suggesting that the large majority of the Voyager magnetometer signature was due to the initial, Alfvénic disturbance. The estimated $7\cdot 10^9$ W is insufficient to power the L-burst component of the decametric radio emissions.

These electron beams may have significant effects on Io itself, and are consistent with observations made by Galileo. The field lines in question connect to Io, so the electron beams generated by the double layers would impact Io's polar regions. For a current density of 10^{-7} A/m^2, and assuming the current is carried by energetic electrons, the energetic electron flux would be $6.25\cdot 10^{11}$ particles/m^2/s. The maximum electron energy would be 98 kV, while average energy would be less by whatever amount had been lost to wave-particle interactions near the double layers. If no energy were lost, and all the particles retained a 98 keV energy, the energy flux from the beam would be 9.8 ergs/s/cm^2, or $7\cdot 10^9$ W in total. The poles of Io, especially on the downstream side where a steady interaction is more likely, would be bombarded by a moderately high flux of energetic electrons. These electron beams would also be present, with similar properties, immediately downstream of Io. So long as the plasma has not been re-accelerated significantly, the potential across the wake would be the same as that across Io.

During Galileo's Dec. 7, 1995 encounter with Io, the Energetic Particle Detector observed bi-directional electron beams 900 km downstream of Io. The particle energies

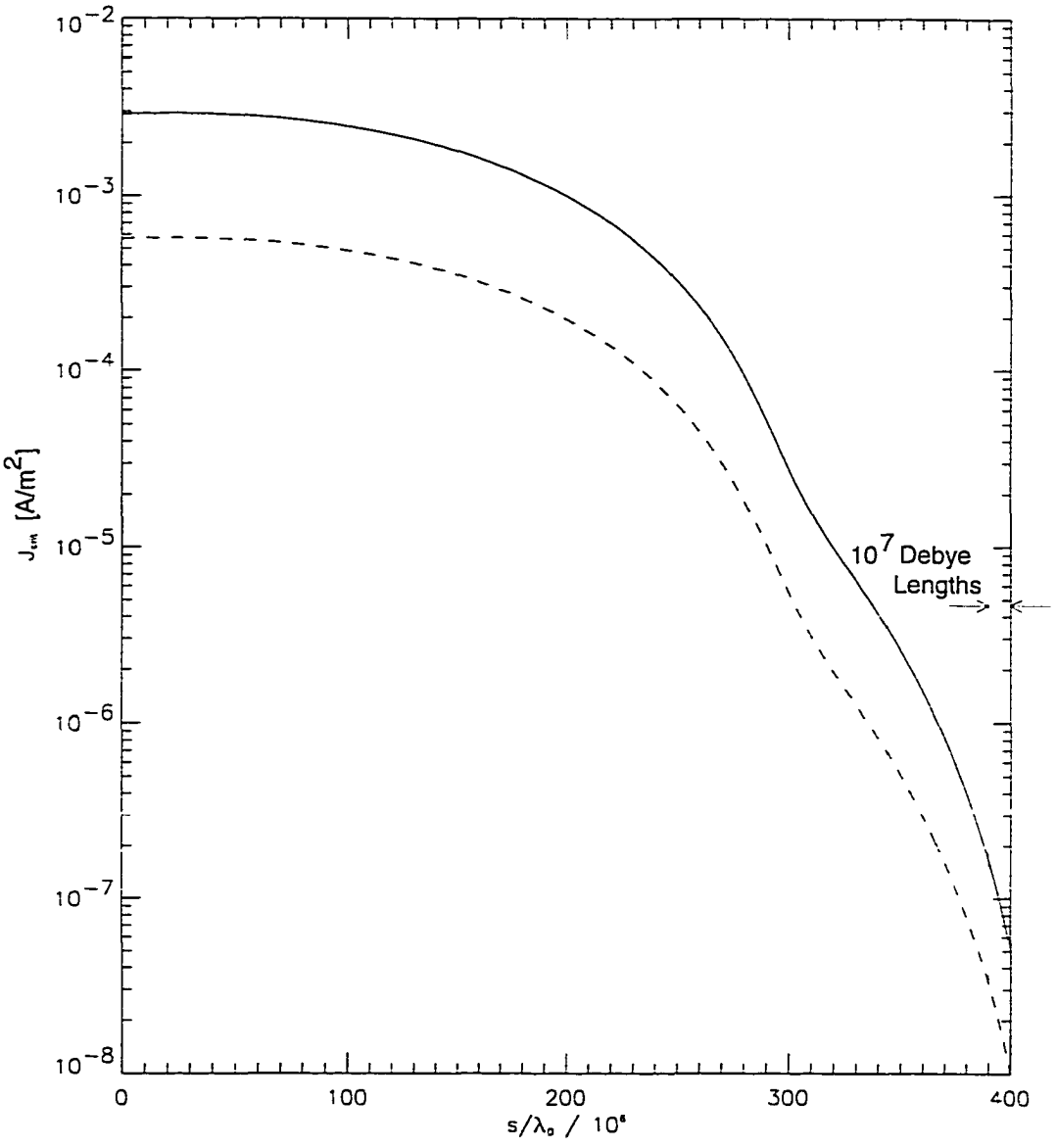


Figure 3.5: The critical current for driving the Buneman instability and causing double layers, as a function of distance along the Io flux tube. The solid line is for $T_e + T_i = 65$ eV, the dashed line for 10 eV.

extended up to ~150 keV but were not observed in the 150 - 200 keV detectors. The energy flux between 20 and 140 keV was 0.05 $ergs/cm^2/s$ and the estimated, total power was 10^9 W [Williams et al., 1996]. Since the signal was strongest at the lowest energy channel, there is substantial uncertainty in the flux and energy of electrons below 15 keV. In any case, the observed particle and energy fluxes are within an order of magnitude of those predicted by my model of double layers, are in the predicted location, and the maximum energy is in reasonable agreement. The agreement in energy and particle flux could be significantly improved by future work: I selected $J_{max} = 10^{-7} A/m^2$ somewhat arbitrarily and based on estimated high-latitude conditions. No effort was made to select a value that would agree with the EPD observations.

3.5 Double layers along the Io wake flux tubes

Double layers forming along the Io flux tube cannot explain the power of the decametric radio emissions. Moreover, auroral UV emissions require an extended particle acceleration along Io's wake. I apply my model of current limiting double layers to the wake. Close to Io, the flow velocity of the wake is nearly stagnant [Frank et al., 1996]; currents coupling the plasma to the ionosphere of Jupiter will eventually accelerate it back up to corotation. I will show that these currents would also produce double layers, and that these double layers would produce both the electron acceleration required by radio emissions, and the long auroral feature observed in the UV.

Since the wake is re-accelerated by the currents closing across it, equation 3.25 must be modified. From the equation of motion,

$$\rho \frac{\partial}{\partial t} v = B_z J_\perp \tag{3.35}$$

and since $\Phi_{wake} \approx E_\perp R_{wake} = -v B_z R_{wake}$,

$$J_\perp = \frac{-\rho}{B_z} \frac{\partial}{\partial t} \frac{\Phi}{R_{wake} B_z} \tag{3.36}$$

I assume that the plasma's velocity is constant along a flux tube, and that ρ and B_z are independent of time. This is equivalent to neglecting transients and nonlinear processes.

Equation 3.36 can then be integrated vertically, and rewritten to give the current per unit length,

$$\begin{aligned} j_{wake} &= -\left(\frac{1}{R_{wake}\mu_0}\int\frac{dz}{V_A^2}\right)\frac{\partial}{\partial t}\Phi_{wake} \\ &\equiv -\frac{C_A}{R_{wake}}\frac{\partial}{\partial t}\Phi_{wake} \end{aligned} \tag{3.37}$$

where C_A is the Alfvén capacitance [Crary and Bagenal, 1997], which has a value of 5 to 6 kF. Assuming no double layers, and combining this with equations 3.26 and 3.27,

$$\begin{aligned} &\frac{\partial}{\partial t}\left(E_{cr}R_{wake}-\Phi\right) = -\frac{2\Sigma_J\sqrt{4-3/L}}{C_A}\left(E_{cr}R_{wake}-\Phi\right) \equiv \frac{-1}{\tau}\left(E_{cr}R_{wake}-\Phi\right) \\ &J_{\parallel} = \frac{2\Sigma_J\sqrt{4-3/L}}{dR_{wake}}\left(E_{cr}R_{wake}-\Phi\right) \end{aligned} \tag{3.38}$$

For typical parameters, τ, the characteristic time for wake re-acceleration is approximately 1300 to 1600 seconds. The velocity and position of a wake flux tube would be

$$\begin{aligned} v &= v_{cr}\left(1-e^{-t/\tau}\right) \\ x &= v_{cr}\tau\left(t-1+e^{-t/\tau}\right) \end{aligned} \tag{3.39}$$

The total energy dissipated by re-accelerating the wake can also be calculated, by integrating the current per unit length times the potential.

$$\begin{aligned} P &= \int_0^\infty dx\, j_{wake}\Phi_{wake} \\ &= -\frac{C_A}{R_{wake}}\int_0^\infty dx\, dt\Phi_{wake}\Phi_{wake} \\ &= \frac{C_A}{R_{wake}}\int_0^{\Phi_0} v\Phi d\Phi \\ &= \frac{C_A}{R_{wake}^2 B_z}\int_0^{\Phi_0}\Phi^2 d\Phi \\ &= \frac{1}{3}C_A\Phi_0^2\frac{v_{cr}}{R_{wake}} \\ &\sim 5\cdot 10^{12}W \end{aligned} \tag{3.40}$$

for $C_A = 5$ kF and $R_{wake} = 1$ R_{Io}.

But equation 3.38 also predicts current densities of approximately $2.5\cdot 10^{-6}A/m^2$, which would be strong enough to generate double layers. Following the model of currents

limited by double layers, the wake would initially re-accelerate as

$$\frac{\partial v}{\partial t} = \frac{J_{max} d}{\int \rho dz / B_z} \equiv a_{wake} \tag{3.41}$$

Assuming $J_{max} = 10^{-7} A/m^2$, $a_{wake} \sim 3.5 m/s^2$. The re-acceleration would continue to be current limited until

$$\begin{aligned} \frac{2\Sigma_J \sqrt{4-3/L}}{dR_{wake}} (E_{cr} R_{wake} - \Phi) &< J_{max} \\ (E_{cr} R_{wake} - \Phi) &< \frac{J_{max} dR_w}{2\Sigma_J \sqrt{4-3/L}} \end{aligned} \tag{3.42}$$

by which time the wake will have a velocity,

$$v = v_{cr} - \frac{J_{max} d}{2 B_0 \Sigma_J \sqrt{4 - 3/L}} \tag{3.43}$$

For typical parameters, this is within 4 km/s of corotation. For all practical purposes, the double layers' maximum current controls the acceleration of the wake. Figure 3.6 shows the wake velocity as a function of position. The effect of double layers is to significantly slow re-acceleration and produce an extended region of particle acceleration. Accelerating the wake from a velocity near zero to 54 km/s, at a rate of 3.5 m/s^2 implies a wake extending for 4x10^8 m or 5.6 R_J.

Near Io, the accelerated electrons would have a spectrum similar to that observed by Galileo. Farther downstream from Io, the peak energy of the particles would decrease, as the potential of the double layers decreased. Since the wake acceleration is linear, the peak energy should scale as $\Phi_0 \left(1 - \sqrt{x/s}\right)$ where s is the length of the wake. The power dissipated by the double layers would be

$$\begin{aligned} P &= \int_0 s J_{max} d\Phi_{DL} dx \\ &= J_{max} d\Phi_0 \int_0 s \left(1 - \tfrac{x}{s}\right) dx \\ &= \tfrac{1}{2} s d J_{max} \Phi_0 \\ &\sim 9 \cdot 10^{11} W \end{aligned} \tag{3.44}$$

which is roughly 20% of the energy dissipated by wake re-acceleration. The remainder of the power goes into Joule heating of Jupiter's auroral ionosphere. The estimated power of the decametric radio emissions is approximately $5 \cdot 10^9$ W. Energetic electrons accelerated

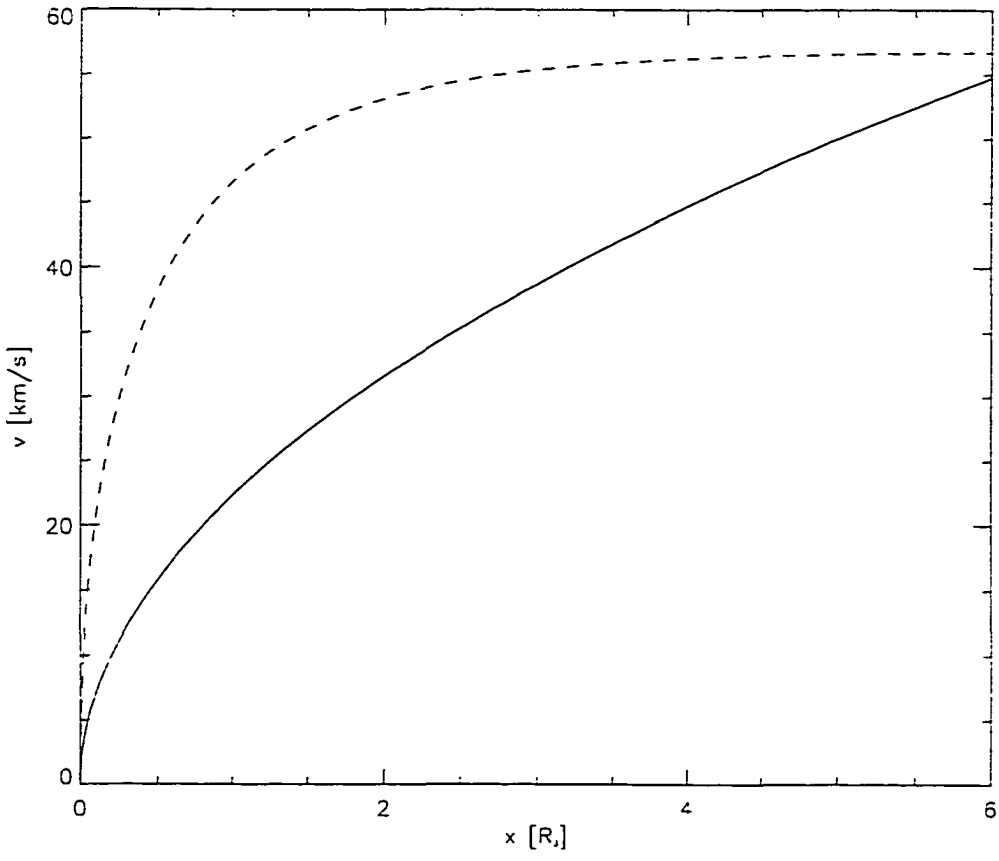


Figure 3.6: The velocity profile of the wake. Assuming double layers with a maximum current of $10^{-7}A/m^2$, solid line, and no double layers, dashed line.

by double layers along Io's flux tube easily have enough power to generate the decametric emissions.

3.6 Summary

In this chapter I have examined the acceleration of electrons by a steady potential drop. The Io interaction would probably evolve into such a steady state on the downstream side of Io, and the re-acceleration of Io's wake would be steady. I have considered field-aligned potential drops in the collisional ionosphere of Jupiter, and found this to be negligible. In a strongly magnetized ionosphere, closure of magnetospheric currents does not result in a significant, field-aligned potential drop.

I discussed the properties of double layers, and showed that a magnetospheric double layer cannot alter its environment on time scales relevant to double layer formation. As such, the flux of particles approaching the double layer should be considered a fixed boundary condition. This is very different from double layers in laboratory experiments and past, theoretical analysis of double layers. Using this assumption, and fluxes at the boundaries appropriate for a background, Maxwellian plasma, I calculated the theoretical structure of a double layer as a function of the field-aligned potential drop. Above a certain potential, 10.9 T_e/e in the case of the boundary conditions I used, the double layer bifurcates. My interpretation is that very strong double layers are unlikely in a magnetosphere, despite having been observed in laboratory experiments, but that many double layers of moderate strength ($\Phi \sim T_e/e$) could form if field-aligned currents drive the Buneman instability.

Applying this model of multiple double layers to the Io interaction, I find that the double layers will begin to form along Io's flux tube if the steady, field aligned current density exceeds a few times 10^{-8} A/m^2, when mapped back Io. For currents over roughly 10^{-7} A/m^2, the current would drive the Buneman instability along approximately 10 million Debye lengths of the flux tube. This region would be sufficiently long for multiple, moderate double layers to cause a potential drop of a few hundred kV, i.e. a potential on par with the induced potential across Io.

To approximate the effects of these double layers, I assume that they would either not be present, or produce what ever potential drop was required to keep the current density

below 10^{-7} A/m^2. From this, I show that the steady currents along the Io flux tube are weak and do not dissipate a significant amount of power. This is of particular interest to interpretations of the Voyager 1 observations of the Io flux tube. I have shown that the steady component of the Io interaction is almost two orders of magnitude weaker than the transient, Alfvénic component of the interaction. Therefore, the Voyager data represents a measurement of the Alfvénic portion of the interaction. In addition, my model of double layers results in the acceleration of electrons along the Io flux tube. The particle energy and energy flux predicted by my analysis is in rough agreement with the bi-directional electron beams observed immediately downstream of Io by the Galileo spacecraft.

Finally, I considered the re-acceleration of Io's wake. With or without the formation of double layers, this process would dissipate roughly 5×10^{12} W. I find that the currents re-accelerating the wake cause double layers, which persist until the wake has been accelerated back up to $\sim 90\%$ of corotation velocity. In the process, the double layers accelerate electrons. The power going into particle acceleration is approximately 10^{12} W. This acceleration process represents enough power, in the form of energetic electrons, to be the source of the bulk of Jupiter's Io-related decametric emissions. In addition, particle precipitation results in the auroral arc extending away from the foot of the Io flux tube. Since the presence of double layers limits the current re-accelerating the wake, the arc signature is longer than would be expected without double layers. Observations of Jupiter's aurora show an unexpectedly long arc, consistent with my model of double layers along the wake flux tube.

Chapter 4

Ion cyclotron waves near Io

4.1 Introduction

In this chapter, I use observations of SO_2^+ ion cyclotron waves near Io, and the absence of S^+ and O^+ waves, to estimate the properties of Io's neutral exosphere and the production rate of ions. These waves are generated by recently ionized pick-up ions, and their spectrum contains information on the the density of these pick-up ions. The amplitude of waves near the SO_2^+ ion cyclotron frequency can be related to the density of pick-up SO_2^+ ions. The absence of S^+ and O^+ indicates that the density of pick-up S^+ and O^+ is sufficiently low, and the thermal background of these ions prevents wave growth. The density of pick-up ions and the ionization rate are important factors for models of the Io interaction, since ionizations slow and heat the ambient plasma. When combined with estimated ionization rates for neutrals, the density of pick-up ions can be used to infer the properties of Io's neutral exosphere. First, I describe the observations and existing analysis of them. Pick-up ion evolution is frequently modeled by assuming that the ions diffuse in velocity space, and that the diffusion coefficient is a tensor which is calculated from quasi-linear theory and which depends on the amplitude of resonant waves (i.e. waves with $\omega = \Omega_c - v_{\|} k$). I summarize the basis for, and limits of, resonant and quasi-linear diffusion. I conclude that these techniques cannot be applied, due to the high amplitude of the observed waves. I then use a warm plasma dispersion relation to estimate marginal stability criteria. Next, I use numerical integrations of test particles to estimate the degree to which S^+ and O^+ ions would be scattered by the observed waves. I use this and marginal stability criteria to place an upper limit on the pick-up densities. Finally, I use these densities and neutral ionization rates to estimate the density and composition of Io's neutral exosphere.

On December 7, 1995, the Galileo spacecraft passed through Io's exosphere and the magnetometer observed intense ion cyclotron waves [Kivelson et al., 1996]. Figure 4.1, from Warnecke et al. [1997] shows the properties of these waves. The perturbation magnetic field is perpendicular to background field (to with in 7°), the waves are left circularly polarized and have frequencies which are lower than the ion cyclotron frequencies of the main ions in the plasma. The ion cyclotron wave is the only mode with these properties. Ion cyclotron waves are frequently observed near a source of plasma, e.g. near comets [Terasawa, 1989]. When the neutrals velocity differs from that of the plasma, newly ionized particles will form a

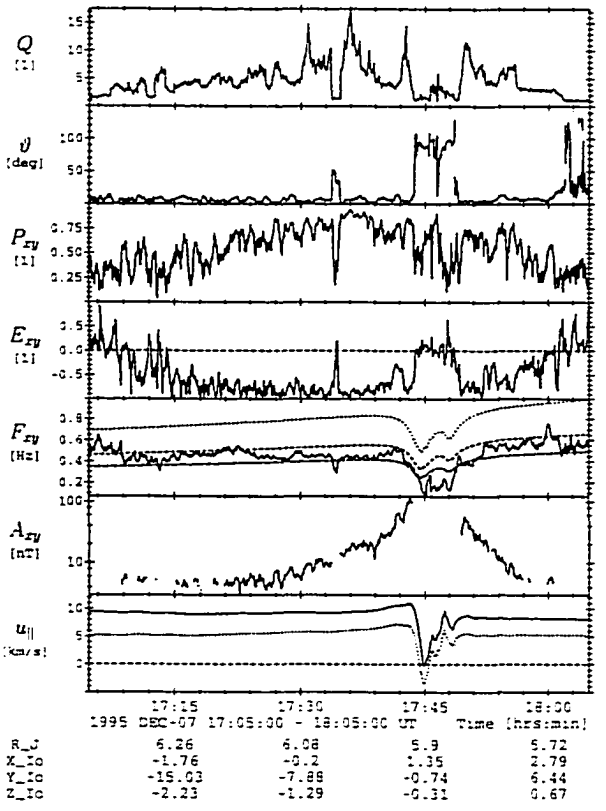


Figure 4.1: Properties of the ion cyclotron waves observed by Galileo in Io's exosphere, from Warnecke et al. [1997]. From top to bottom: Q, the ratio of the intermediate to minimum eigenvalues of the covariance matrix, a measure of the how "wave-like" the observed magnetic field was; ϑ the angle between the background field and the minimum variance direction (i.e. the direction of $\vec{k}$; P, the degree of polarization, E the ellipticity; F, the peak frequency, along with the S^+,SO^+ and SO_2^+ gyrofrequencies; A, the amplitude; and u, the field-aligned component of velocity for pick-up ions (solid line) and the Galileo spacecraft(dotted line).

ring-beam distribution. Such distributions are generally unstable and generate ion-cyclotron waves in the process of diffusing into a more stable distribution.

The ion cyclotron waves observed by Galileo were very strong, with peak amplitudes of almost 100 nT or slightly more than 6% of the background field strength, and had peak frequencies near the SO_2^+ gyrofrequency. Spectra of the waves, from Warnecke et al. [1997] are shown in figure 4.2. No waves near the S^+ or O^+ cyclotron frequencies were observed, despite the fact that these ions are more common in the Io plasma torus and have longer lifetimes. SO_2^+ waves were observed out to distance of roughly 10 Io radii from Io in the anti-jovian direction, and 6 Io radii towards Jupiter. An exosphere can only be gravitationally bound to Io at distances less than the Hill radius, which is 5.8 R_{Io} in the case of Io. The source of pick-up ions causing the waves seen at greater distances must be due to escaping, rather than gravitationally-bound neutrals. The asymmetry in the wave amplitude also suggests that the escape process is asymmetric and preferentially directed in the anti-jovian direction. Wilson, 1996, proposed two theoretical mechanisms which cause asymmetric escape in this direction, both resulting from the induced, corotation electric field (which is in the anti-jovian direction.)

The Voyager spacecraft during its 1979 encounter with Io did not observe any ion cyclotron waves, down to a few nT noise level. Its trajectory passed very close to Io's flux tube but 11 Io radii south of Io itself. If ion cyclotron waves similar to those Galileo observed were present and able to freely propagate, the Voyager magnetometer should have detected them. As I show in this chapter, these waves are strongly damped by non-resonant scattering of S^+ ions and would be attenuated after propagating 11 Io radii along the field lines. This may account for the lack of a Voyager detection.

Huddleston et al. [1997] performed a warm plasma analysis of these waves. They assumed that the distribution of each species is

$$f(v_{\|}, v_{\perp}) \propto \exp\left[-\frac{(v_{\|}-v_b)^2}{a_{\|}^2} - \frac{(v_{\perp}-v_r)^2}{a_{\perp}^2}\right] \tag{4.1}$$

where $a_{\|}$ and $a_{\perp}$ are the parallel and perpendicular thermal velocities, v_b a parallel beam velocity and v_r the ring velocity of a ring-beam distribution. This distribution is illustrated in figure 4.3. For this distribution, the dispersion relation for parallel propagating ion cyclotron

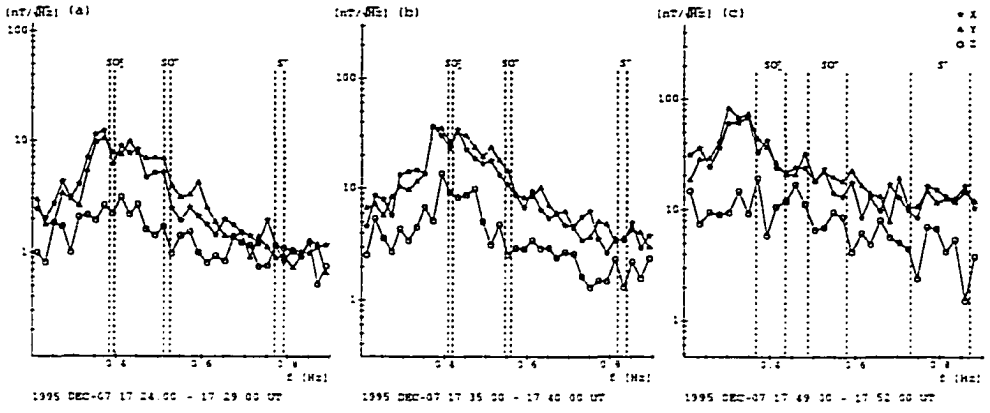


Figure 4.2: Spectra of ion cyclotron waves, from Warnecke et al. [1997].

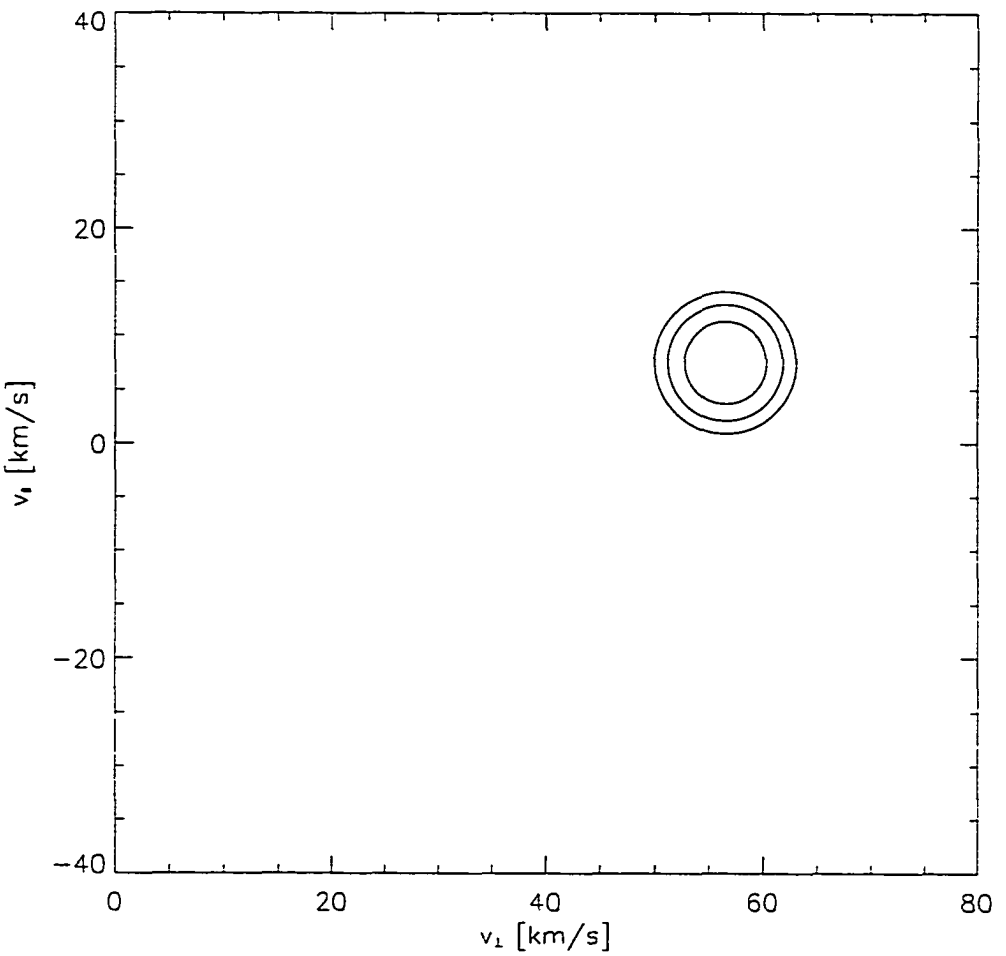


Figure 4.3: Contours of a ring-beam (pick-up) distribution, for $a_{\|}$=$a_{\perp}$ = 2.5 km/s, v_b = 7.5 km/s, v_r = 56.5 km/s, consistent with a 5 eV plasma and field line tilt at the time of the Galileo encounter. Contour levels are of 0.1, 0.01 and 0.001 times the peak phase space density.

waves in a warm plasma is

$$0 = c^2k^2 - \omega^2 - \sum_{species} \omega_{ps}^2 \{\xi_0 Z(\xi) + (A-1)[1 + \xi Z(\xi)]\} \tag{4.2}$$

where $\xi = (\omega - kv_b - \Omega_s)/ka_{\|}$, $\xi_0 = (\omega - kv_b)/ka_{\|}$, $\Omega_s = q_s B/m_s$ is the cyclotron frequency and $\omega_s = \sqrt{q_s^2 n_s/\epsilon_0 m_s}$ is the plasma frequency of a species, and $Z(\xi)$ is the plasma dispersion function of Fried and Conte [1961]. A, the thermal anisotropy, is $T_\perp/T_{\|}$ for the above distribution. However, this can be generalized. For the distribution,

$$f(v_{\|}, v_\perp) \propto g(v_\perp)\exp\left[-\frac{(v_{\|} - v_b)^2}{a_{\|}^2}\right] \tag{4.3}$$

where $g(v_\perp)$ is arbitrary, then equation 4.2 is valid, and

$$A = < v_\perp^2 > /a_{\|}^2 = \frac{2\pi}{a_{\|}^2}\int_0^\infty v_\perp^3 g(v_\perp) dv_\perp \tag{4.4}$$

[Huddleston et al., 1998]. I will use this generalized anisotropy in this chapter, since it may be easily calculated from the results of my test particle integrations. I will also assume that the parallel distribution of my test particles is approximately Maxwellian, and adopt equation 4.2, setting $a_{\|}^2 = < v_{\|}^2 >$.

One result of Huddleston et al.'s analysis explains the lack of S^+ and O^+ waves. While pick-up ions cause an instability and wave growth, a thermal core population of the same species causes damping and prevent the instability. Since there is a significant core population of S^+ and O^+ ions in the plasma torus, the presence or absence of waves at these ions' gyrofrequencies depends on the relative abundance of pick-up ions. I use this property to set upper limits on the density of S^+ and O^+ pick-up ions. In contrast, SO_2^+ has a short life time in the torus, of order one hour, before being dissociated by electron impacts. As a result, there would be no significant population of thermalized SO_2^+ and even a weak source of SO_2^+ pick-up ions would result in ion cyclotron waves.

4.2 Ion scattering and Quasi-linear diffusion

Quasi-linear diffusion is based on the second order evolution of a spatially averaged distribution function. The distribution function is separated into two components

$$f(\vec{s},\vec{v},t) = f_0(\vec{v},t) + f_1(\vec{v})e^{i(\vec{k}\cdot\vec{s}-\omega t)} \tag{4.5}$$

where f_1 is an oscillation due to plasma waves, and where I have simplified the problem by including only one oscillatory mode rather than a spectrum of waves. Provided that

$$\frac{\partial f_0}{\partial t} \ , \ \frac{q}{m}(\vec{E}+\vec{v}\cdot\vec{B})\vec{\nabla}_v f_1 \ll (\omega - \vec{v}\cdot\vec{k})f_1 \ , \ \frac{q}{m}(\vec{E}+\vec{v}\cdot\vec{B})\vec{\nabla}_v f_0 \tag{4.6}$$

the Vlasov equation can be separated into

$$\begin{aligned} f_1 &= -\tfrac{1}{\omega-\vec{v}\cdot\vec{k}}\tfrac{q}{m}(\vec{E}+\vec{v}\cdot\vec{B})\vec{\nabla}_v f_0 \\ \tfrac{\partial f_0}{\partial t} &= -\tfrac{q}{m}(\vec{E}+\vec{v}\cdot\vec{B})\vec{\nabla}_v f_1 \end{aligned} \tag{4.7}$$

With some manipulation, this can be reduced to a diffusion equation,

$$\frac{\partial f_0}{\partial t} = \vec{\nabla}_v\left(\overleftrightarrow{D}\ \vec{\nabla}_v f_0\right) \tag{4.8}$$

where $\overleftrightarrow{D}$ is a tensor diffusion coefficient, which depends on the amplitude and phase velocity of the wave. The magnitude of the parallel component, $D_{\mu,\mu}$, expressed in terms of pitch angle, is

$$D_{\mu,\mu} = \frac{\pi}{2}\frac{\Omega^2}{B^2}\frac{1-\mu^2}{v|\mu|}I\left(\frac{\Omega-\omega}{v_z}\right) \tag{4.9}$$

where $I(k)$ is the spectral energy density, $< \delta B^2 > \equiv \int_{-\infty}^{\infty} I(k)dk$, [Lee, 1982]. Rewriting this in terms of parallel velocity for an ion with $v_z = 0$ and converting I to I_ω, a spectral energy density as a function of ω rather than k, this becomes

$$D_{||,||} = \frac{\sqrt{\pi}}{2}v_g v_\perp \frac{\Omega^2}{B^2}I_\omega(\Omega-\omega) \tag{4.10}$$

In the case of the ion cyclotron waves observed near Io, I_ω was as high as 10^4 nT^2/Hz near the SO_2^+ gyrofrequency and 100 nT^2/Hz near the S+ gyrofrequency. Inserting this into

equation 4.10, I obtain $D_{\|,\|} \sim 2 \cdot 10^7 \ \ m^2/s^3$ for SO_2^+ and $D_{\|,\|} \sim 1 \cdot 10^6 \ \ m^2/s^3$ for S+. Estimating $f_1(\omega - vk)$ from the Vlasov equation, I find that the ordering assumptions of quasi-linear diffusion require

$$D_{\|,\|} \ll \frac{q\omega\sqrt{<\delta B^2>}}{m} a \tag{4.11}$$

where a is a typical thermal velocity of the ions. For 5 eV pick-up ions near Io, this requires a diffusion coefficient well under $5 \cdot 10^7$ for SO_2^+ and $1 \cdot 10^7$ for S^+. As a result, quasi-linear diffusion cannot be used to model the evolution of SO_2^+ pick-up ions near Io, but may be reasonably accurate for the S^+ ions. Since the wave amplitude is zero near the oxygen gyrofrequency, quasi-linear theory predicts no diffusion of oxygen ions and $D_{\|,\|} = 0$ automatically satisfies the assumptions of quasi-linear theory.

Equation 4.10 also assumes resonant diffusion, and its derivation explicitly ignores non-resonant effects [Kennel and Engelmann, 1966; Shapiro and Shevchenko, 1962] I will use an example of single particle motion to illustrate the justification behind and limits to resonant diffusion. If an ion is accelerated by an ion cyclotron wave with frequency ω and $\vec{k} \| \vec{B}_0$, its equations of motion are

$$\begin{aligned} \tfrac{\partial \vec{v}}{\partial t} &= \tfrac{q}{m}\left[\vec{E} + \vec{v} \times \left(\vec{b} + \vec{B}_0\hat{z}\right)\right] \\ \vec{E} &= \tfrac{E}{\sqrt{2}}\left[\cos(kz - \omega t + \phi)\hat{x} - \sin(kz - \omega t + \phi)\hat{y}\right] \\ \vec{b} &= \tfrac{kE}{\omega\sqrt{2}}\left[\sin(kz - \omega t + \phi)\hat{x} + \cos(kz - \omega t + \phi)\hat{y}\right] \end{aligned} \tag{4.12}$$

By the substitution

$$\begin{aligned} v_x &= -v_\perp \sin(\Omega t) \\ v_y &= -v_\perp \cos(\Omega t) \end{aligned} \tag{4.13}$$

and assuming v_z is approximately constant so that $z \sim v_z t$ and integrating, this becomes

$$v_\perp \sim \frac{qE}{m\sqrt{2}}\left(1 - \frac{v_z k}{\omega}\right)\frac{\cos(\phi) - \cos\left[\phi + (\Omega + kv_z - \omega)\,t\right]}{(\Omega + kv_z - \omega)\,t}\ t \tag{4.14}$$

For $(\Omega + kv_z - \omega)\,t \ll 2\pi$, the particle is accelerated by the wave, regardless of resonance conditions. However, as time passes, the average change in the particle's velocity will go to zero unless $(\Omega + kv_z - \omega) = 0$. Non-resonant scattering may, therefore, be neglected if a characteristic time for scattering is sufficiently long, i.e. $(\Omega + kv_z - \omega)\,\tau_{scattering} \gg 2\pi$. For

S^+ pick-up ions near Io, a characteristic diffusion time would be of order 1 second. As a result, non-resonant scattering would be significant, and the ions would be affected by waves satisfying

$$|\Omega + kv_z - \omega| \lesssim 2\pi\ s^{-1} \tag{4.15}$$

rather than $\Omega + kv_z - \omega = 0$. This frequency range includes all of the observed ion cyclotron waves. Therefore, the evolution of the S^+ pick-up ions cannot be accurately modeled without including non-resonant scattering.

Frequently models of pick-up ion scattering [Huddleston et al., 1997; Terasawa, 1989] take advantage of the conserved quantity,

$$v_\perp^2 + \left(v_\parallel - \frac{\omega}{k}\right)^2 = \text{constant} \tag{4.16}$$

and assume the ions diffuse into a bispherical distribution. This conserved quantity is a result of Faraday's equation and the transformation of electromagnetic fields into a moving coordinate system. In the coordinate system $\vec{v}' = \vec{v} - \omega/k\hat{k}$, the wave fields are static and

$$\vec{E}' = \frac{1}{\sqrt{n^2-1}}\frac{\omega}{k}\left(\hat{k}\times\vec{B}_0\right) - \frac{n}{(n^2-1)\left(\sqrt{n^2-1}+n\right)}\left(\vec{k}\cdot\vec{E}\right)\hat{k} \tag{4.17}$$

where $n = ck/\omega$ is the wave's index of refraction. If the waves are purely electromagnetic waves, $\vec{k}\cdot\vec{E} = 0$, and propagate parallel to the background magnetic field $\hat{k}\times\vec{B}_0 = 0$, then the electric field in the wave frame is zero and the particle's energy in this frame will be constant. An additional requirement is that the waves are not dispersive, since a phase velocity, ω/k, is assumed. If the phase velocity also varies, then the "energy of the ion in the wave's reference frame" is ambiguous.

The requirement for a single phase velocity is a particular problem, since particles in a ring distribution generate waves propagating parallel and anti-parallel to the magnetic field. Even if the waves were not dispersive, there would still be $v_\phi = \omega/k$ and $v_\phi = -\omega/k$ waves. This results in diffusion away from the wave characteristics. In the case of cometary interactions, this can generally be considered a second order effect [Terasawa, 1989]. In that case, the waves have a low phase velocity, $\omega/k \ll v_{pu}$, and the characteristics of the parallel and anti-parallel waves are nearly the same. The result is rapid scattering in pitch angle and

a much slower diffusion in velocity. Figure 4.4 shows a sketch of this, based on figure 3b of Terasawa [1989]. In the case of the ion cyclotron waves near Io, the phase velocity is of order the pick-up velocity. This causes rapid diffusion, both in pitch angle and velocity. Figure 4.4 also shows a sketch of this, and the contrasting direction of the wave characteristics.

Formally, this can be demonstrated in the limit of quasi-linear diffusion. In that approximation, the diffusion tensor, for the n=1 resonance with a parallel propagating electromagnetic wave, is a scalar times

$$\overleftrightarrow{D}_{+} \propto \vec{a}\vec{a}^{*} = \begin{bmatrix} (\omega/k - v_{\parallel})^2 \;\; \hat{v}_{\perp}\hat{v}_{\perp} & (\omega/k - v_{\parallel})\, v_{\perp} \;\; \hat{v}_{\parallel}\hat{v}_{\perp} \\ (\omega/k - v_{\parallel})\, v_{\perp} \;\; \hat{v}_{\perp}\hat{v}_{\parallel} & v_{\perp}^2 \;\; \hat{v}_{\parallel}\hat{v}_{\parallel} \end{bmatrix} \tag{4.18}$$

where $\hat{v}_{\parallel}$ and $\hat{v}_{\perp}$ are unit vectors in velocity space, parallel and perpendicular to the background magnetic field. The diffusion tensor for an anti-parallel propagating wave of equal amplitude and frequency would be identical, except that k would be replaced by $-k$. An ion with $v_{\parallel} = 0$ would be resonant with both these waves (assuming $\omega = \Omega_c$), and the diffusion tensor would be

$$\overleftrightarrow{D} = \overleftrightarrow{D}_{-} + \overleftrightarrow{D}_{+} \propto \begin{bmatrix} \omega^2/k^2 \;\; \hat{v}_{\perp}\hat{v}_{\perp} & 0 \\ 0 & v_{\perp}^2 \;\; \hat{v}_{\parallel}\hat{v}_{\parallel} \end{bmatrix} \tag{4.19}$$

The ratio of the parallel and perpendicular diffusion coefficients is

$$D_{\parallel\,\parallel}/D_{\perp\,\perp} = v_{\perp}^2/v_{\phi}^2 \tag{4.20}$$

Which is a large number for cometary interactions, but small or of order unity in the case of the Io plasma torus. As a result, cometary pick-up ions would be expected to scatter in pitch angle while Io torus pick-up ions would be expected to scatter in velocity and, to a lesser extent, pitch angle. It is convenient to note that, for heavy pick-up ions (i.e. relative to the average ion mass) with no beam velocity, $v_{\perp}^2/v_{\phi}^2 \approx v_{flow}^2/V_A^2 = M_A^2$. Anisotropic diffusion of this form is also consistent with observations of Saturn's inner magnetosphere, where the thermal anisotropy was observed to be large, $A \sim 6$ and the Alfvén Mach number small, $M_A \approx 0.06$ [Lazarus and McNutt, 1983].

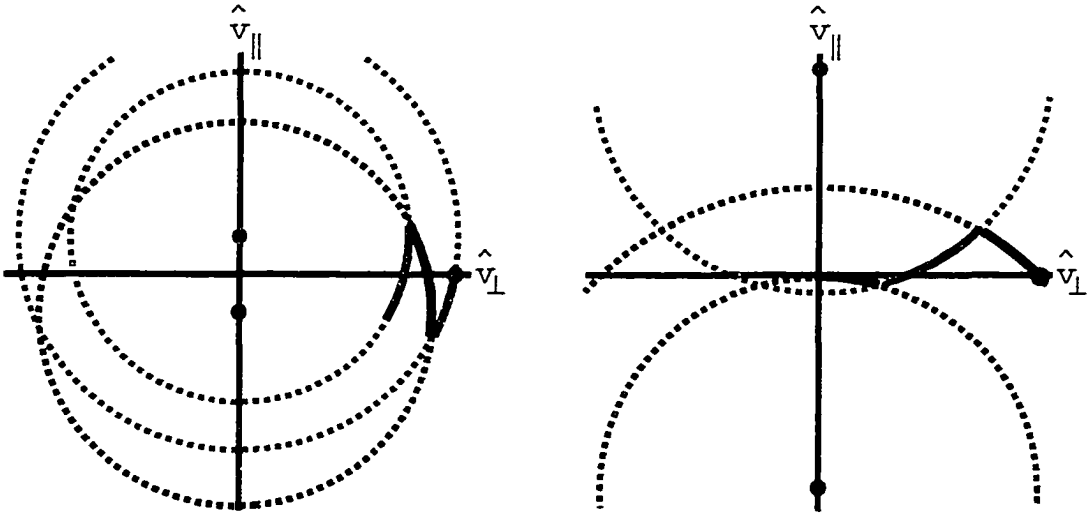


Figure 4.4: A cartoon of particle scattering by two waves. Solid lines show the path a particle might follow through phase space, while dotted lines show wave characteristics, along which the particles are scattered. The parallel propagating wave is shown in red while the anti-parallel in blue. Left, from Terasawa [1989], is the situation when the phase velocity is small compared to the particle velocity; right, the case for a large phase velocity.

4.3 Limits on non-thermal S^+ and O^+ abundance

The absence of a peak in the wave spectrum near the S^+ and O^+ gyrofrequencies indicates that the thermal population of S^+ and O^+ is able to suppress the instability cause by pick-up ions. In this section, I describe what restrictions this places on the density of pick-up ions.

In a bi-Maxwellian plasma, anisotropic ions will cause wave damping at frequencies $\omega > (A-1)/A\Omega_c$, but are unstable and will cause wave growth at lower frequencies [Kennel and Petschek, 1966; Scharer and Trivelpiece, 1967]. For more complex distributions this critical frequency may be altered, or the instability may not exist at all. In the limit, $|(\omega - \Omega_c)/ak| \gg 1$ and $|\gamma| \ll |\omega_r|$ (where $\omega = \omega_r + i\gamma$), the growth rate can be derived from equation 4.2 [Machida et al., 1988]

$$\gamma \approx -\sum_s \frac{\omega_{ps}^2\sqrt{\pi}}{a_{\|s}k}\left(2\omega + \frac{\Omega_{cs}\omega_{ps}^2}{\omega(\omega-\Omega_{cs})^2}\right)^{-1}\left[A\omega - (A-1)\,\Omega_{cs}\right]e^{-(\omega-\Omega_{cs})^2/a_{\|s}^2k^2} \qquad (4.21)$$

where ω_{ps},Ω_{cs}, and $a_{\|s}$ are the plasma frequency, gyro-frequency and parallel, thermal velocity of a species, and the sum is over all species. The contribution from each species is exponentially dependent on $(\omega - \Omega_{cs})^2/a_{\|}^2k^2$. Since $a_{\|} \ll \omega/k$ for all species in the Io torus, I retain only one species and neglect the contribution from the others. If this species has two components, an isotropic core and an anisotropic, hot component, then

$$\gamma \propto \omega\omega_{pc}^2 e^{-(\omega-\Omega_c)^2/a_c^2k^2} + \left[\omega + (A-1)(\omega-\Omega_c)\right]\omega_{ph}^2 e^{-(\omega-\Omega_c)^2/a_h^2k^2} \qquad (4.22)$$

and waves are unstable if

$$\frac{n_h}{n_c}\left[A - (A-1)\,\frac{\Omega_c}{\omega}\right]e^{-(\omega-\Omega_c)^2(1/a_h^2 - 1/a_c^2)/k^2} < -1 \qquad (4.23)$$

Due to the exponential term, an anisotropic, hot component can be stable at all frequencies. The term on the left hand side of equation 4.23 is negative when $\omega < (A-1)/A\Omega_c$ If the isotropic population's parallel temperature is higher than the hot component's, then $1/a_h^2 - 1/a_c^2$ is positive, and the term on the left hand side may be exponentially small when $\omega < (A-1)/A\Omega_c$, and in this case the term would never be less than -1, and the distribution

would be stable. Similarly, when the parallel temperature of the hot component is larger than that of the thermal core, the exponential term in 4.23 is large, and the damping from the core is generally weaker than the instability from the anisotropic component.

In the special case, $a_h = a_c$, the waves are unstable when

$$\omega < \frac{n_h (A-1)}{n_h A + n_c} \Omega_c \tag{4.24}$$

In a multi-species plasma this instability may not occur at all: If the critical frequency is sufficiently low, it may approach the gyrofrequency frequency of another, higher mass species. In that case, additional damping might occur and prevent an instability.

Without making the above assumptions, equation 4.2 may be solved to determine the growth rate as a function of pick-up ion density, parallel temperature and anisotropy. The possible results vary by orders of magnitude, depending on the parameters. To place useful limits on the pick-up ion densities, some constraints on the parallel temperature and anisotropy are necessary.

4.4 Test particle scattering analysis

To reduce the uncertainty in the pick-up ions' temperature and anisotropy, I integrated the trajectories of test particles and calculated the extent to which they would be non-resonantly scattered by the SO_2^+ ion cyclotron waves. The wave spectrum was not calculated self-consistently: I simply adopted the spectrum observed by Galileo. Figure 4.5 shows the spectrum Galileo observed between 17:35-17:40 SCET and the model spectrum I have adopted. The model spectrum is

$$I(\omega) = I_0 \left(\frac{\omega}{\Delta\omega}\right)^N e^{-\omega^2/\Delta\omega^2} + I_{noise} \tag{4.25}$$

with values $\Delta\omega = 0.18$ Hz, $N = 9.5$, $I_{noise} = 3.7$ nT^2/Hz, which are the best fit parameters for a spectrum of this form. I_0 was set so that $\sqrt{< \delta b^2 >} = 0.025 B_0$. The component of the observed spectrum which I attribute to noise, I_{noise}, was not included in the numerical integrations. Figure 4.6 shows the observed and model wave forms.

I have assumed that waves of equal amplitude propagate parallel and anti-parallel

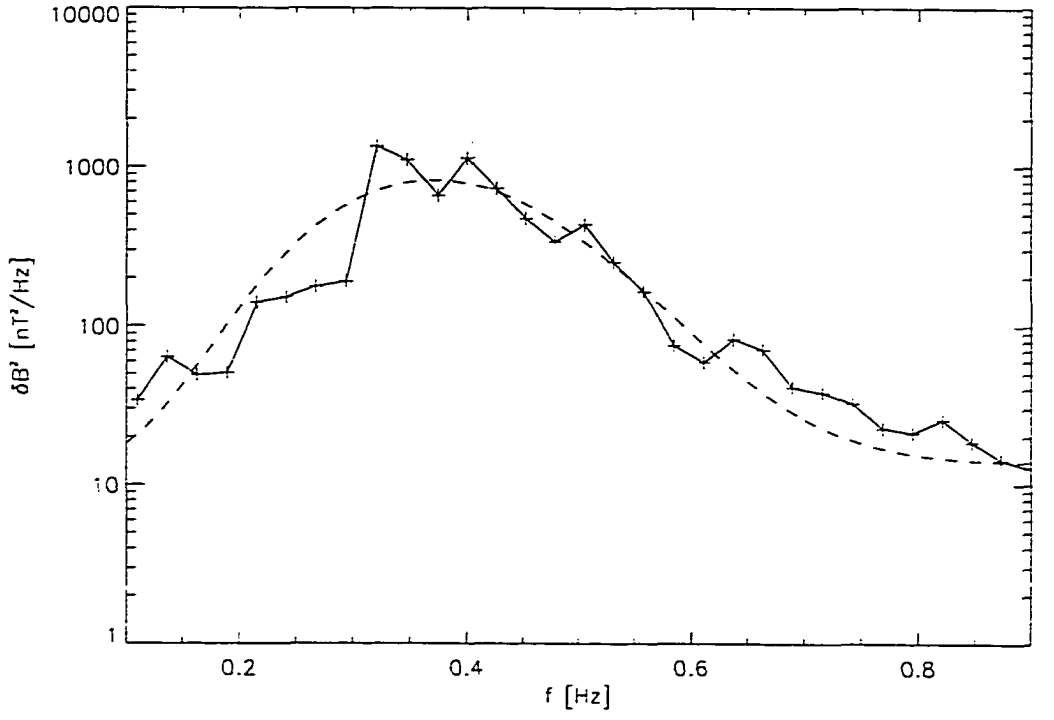


Figure 4.5: The spectrum Galileo observed between 17:35-17:40 SCET, solid line, and the model spectrum used in test particle integrations, dashed lines.

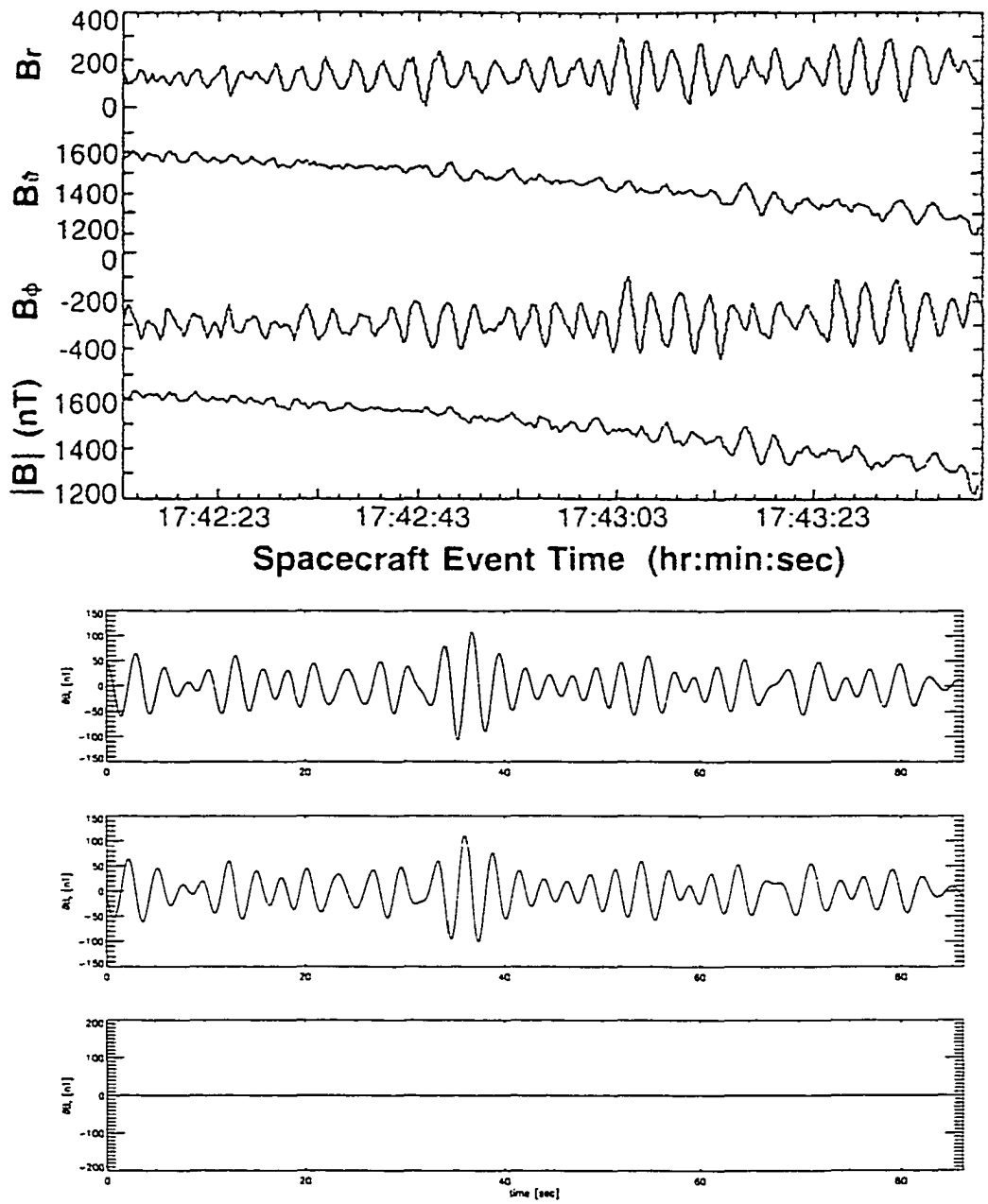


Figure 4.6: Top: wave forms of the ion cyclotron waves observed by Galileo, from Huddleston, 1997. Bottom: Model wave forms used in the test particle integrations.

to the magnetic field, that $k_\perp = 0$, and solved equation 4.2 to obtain k. I used a plasma with five ion populations: Isotropic, thermal S^+, S^{++}, O^+ and O^{++}, with densities and temperatures based on Galileo observations Table 1.1, and ring population of SO_2^+ with a density of 20 cm^{-3}, $T_\parallel = 5$ eV and A = 100. The SO_2^+ parameters imply $T_\perp \sim 0.5 E_{pu.SO_2^+}$. The choices of temperature and density are arbitrary. I have selected 5 eV for consistency with Huddleston et al. (1997), and 20 cm^{-3} is in the range of values suggested by Huddleston et al. (1997) and Warnecke et al. (1997). In addition, $T_\perp \sim 0.5 E_{pu.SO_2^+}$ and my choice of anisotropy are consistent with scattering into a bispheric shell by waves with $v_\phi \gg v_{pu}$. In general, the dispersion relation is insensitive to these values, since SO_2^+ is a minor species and $\omega - \Omega_{c.SO_2^+} \gg a_{SO_2^+} k$ for most of the spectrum. I used a magnetic field strength of 1800 nT for the calculations. Figure 4.7 shows the resulting wave vectors. There is a narrow stop band above the SO_2^+ gyrofrequency and a somewhat wider stop band at the S^+ gyrofrequency. Figure 4.8 shows the phase velocity and figure 4.9 the group velocity.

Using this wave spectrum, I numerically integrated the equations

$$\begin{aligned} \frac{\partial}{\partial t}\vec{s} &= \vec{v} \\ \frac{\partial}{\partial t}\vec{v} &= \frac{q}{m}\left[\vec{E} + \vec{v} \times \left(\vec{B}_0 \hat{z} + \vec{b}\right)\right] \end{aligned} \tag{4.26}$$

I use an adaptive step size, fourth-order Runge-Kutta integrator [Press et al., 1992], with the step size varying to keep the error in velocity (position) near 10^{-9} thermal velocities (Larmor radii) per step. This step size was typically around 0.05 to 0.1 Ω_c^{-1}. Since the results, below, are for integrations of up to 150 Ω_c^{-1} and the errors would be stochastic, the results should be accurate to roughly $5 \cdot 10^{-8}$ times the thermal velocity or the Larmor radius. The integrator itself uses non-dimensional units, based on the S^+ gyrofrequency and thermal velocity and the background field strength.

To test the integrator, I performed integrations of scattering of 1000 particles by a low amplitude, non-dispersive wave. The particles are initially placed in a ring distribution, a Gaussian with a temperature of $0.01a$ centered on $\vec{v} = 2.5a\hat{v}_\perp$. The waves are given a phase velocity of $\omega/k = +2a$. The initial distribution is similar to that of pick-up ions near Io, and the wave velocity is also similar to that of the ion cyclotron waves near Io. The wave spectrum was Gaussian centered on Ω_c and with width of 0.5 Ω_c. In this case, the particles

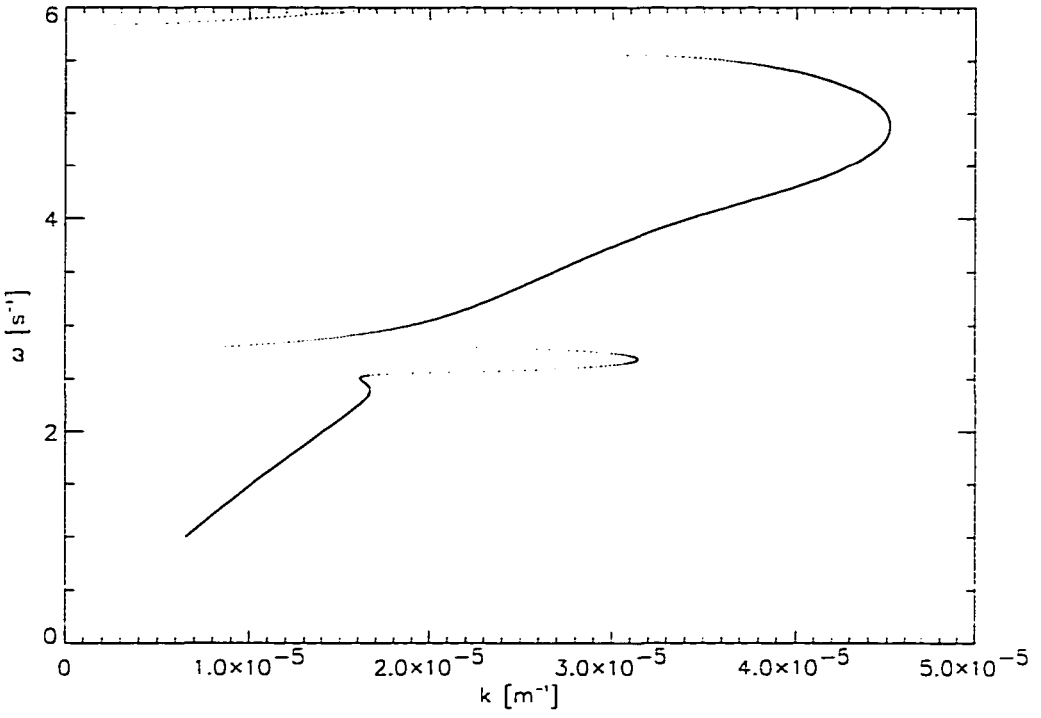


Figure 4.7: The dispersion relation for the ion cyclotron waves used in the test particle integrations.

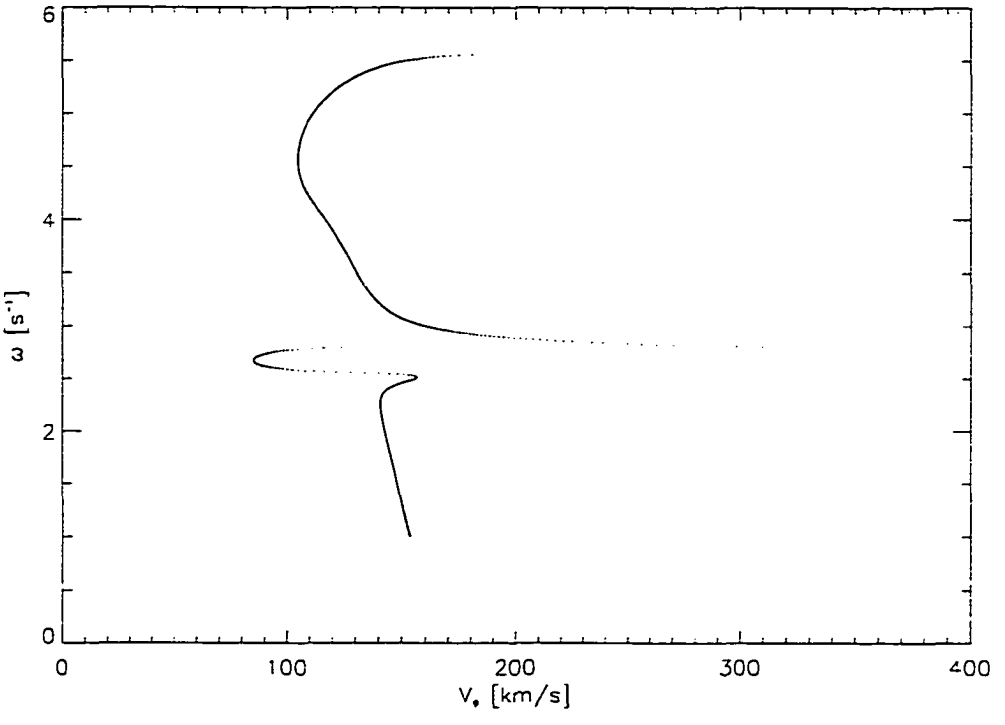


Figure 4.8: The phase velocity for the ion cyclotron waves used in the test particle integrations.

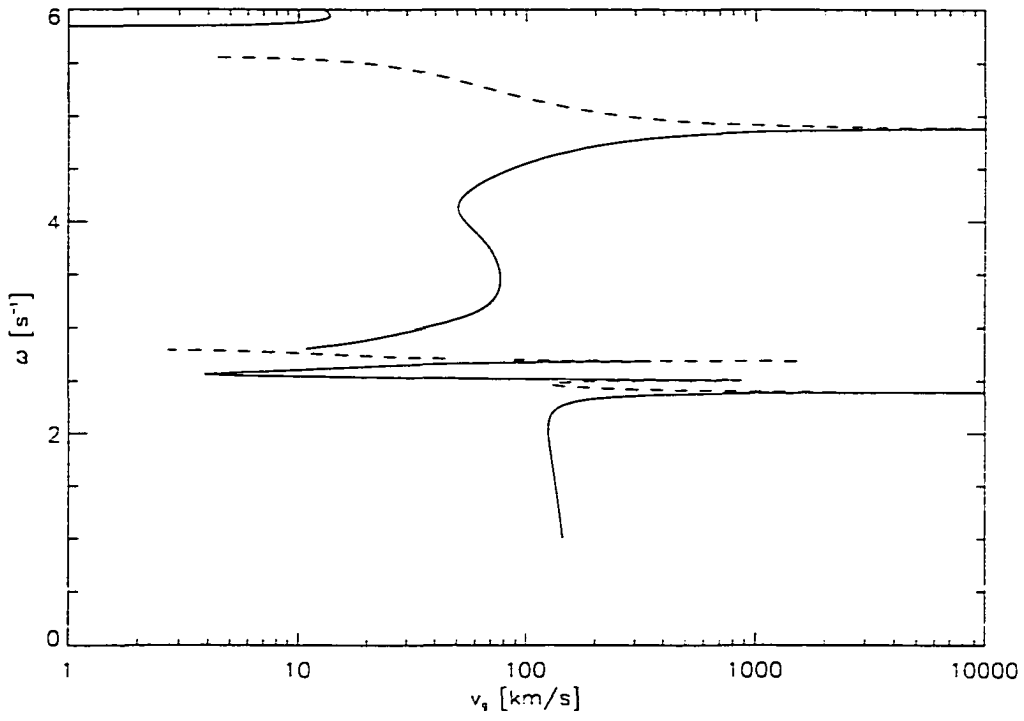


Figure 4.9: The group velocity for the ion cyclotron waves used in the test particle integrations. Dashed lines indicate a group velocity which is anti-parallel to the phase velocity.

should scatter along the characteristic direction, conserving $v_\perp^2 + (v_\parallel - w/k)^2$, and quasi-linear theory should predict their diffusion rate, with $< v^2 >= Dt$. The results of the integrations are shown in the following figures. Figure 4.10 shows the phase space distribution at the end of one integration, figure 4.11 shows the rms velocity of the test particles as a function of time and figure 4.12 shows the rms velocity at the end of each integration, for several wave amplitudes. These results agree with quasi-linear theory, with two exceptions, both of which can be seen in figure 4.11. There is an initial period of rapid scattering and a weak oscillation in $< v^2 >$. The initial acceleration is consistent with linear acceleration for approximately one to two autocorrelation times [Sanbonmatsu et al., 1997] and represents a transient which does not affect the long-term results. The oscillation is statistical, and due to my use of only 1000 particles: When I performed integrations with 250 particles, the amplitude of the oscillations was a factor of two larger, suggesting the $1/\sqrt{N}$ scaling of statistical noise.

For the S^+ and O^+ integrations, I used 2000 ions, 1000 in a isotropic Maxwellian distribution and 1000 in a ring-beam distribution. The relative numbers do not imply anything about the actual ratio of pick-up versus thermal ions. In subsequent results, I will use average properties (temperature, anisotropy, etc) of the populations. 1000 is simply a large enough number to minimize statistical noise while being small enough to require minimal computer time. The ring-beam ions had a temperature of 5 eV centered on the pick-up velocity. Figures 4.13 and 4.14 shows the phase space distribution of the test particles for four times. Figure 4.15 shows the parallel temperature of the two populations and figure 4.16 the anisotropy. The S^+ ions scattered very rapidly. The ring-beam population initially diffused perpendicular to the field into a more isotropic distribution while the core population become hot and anisotropic. Due to the strong ion heating, these integrations cannot be considered accurate after $t \sim 10\Omega^{-1}$. By that time, $\frac{1}{2} n_{S^+} k \left(T_\parallel + 2T_\perp\right) \sim 0.1 \frac{\delta B^2}{2\mu_0}$, and the ion heating would result in significant wave damping. Since I assumed the wave amplitude was constant, the results at later times are not self-consistent. I conclude that S^+ ions, in both the pick-up and core populations, are strongly and non-resonantly scattered and would evolve into a moderately anisotropic distribution, but the longer-term evolution cannot be simulated by test particle integrations.

In contrast, my integrations of oxygen ions show no diffusion at all. The distribution

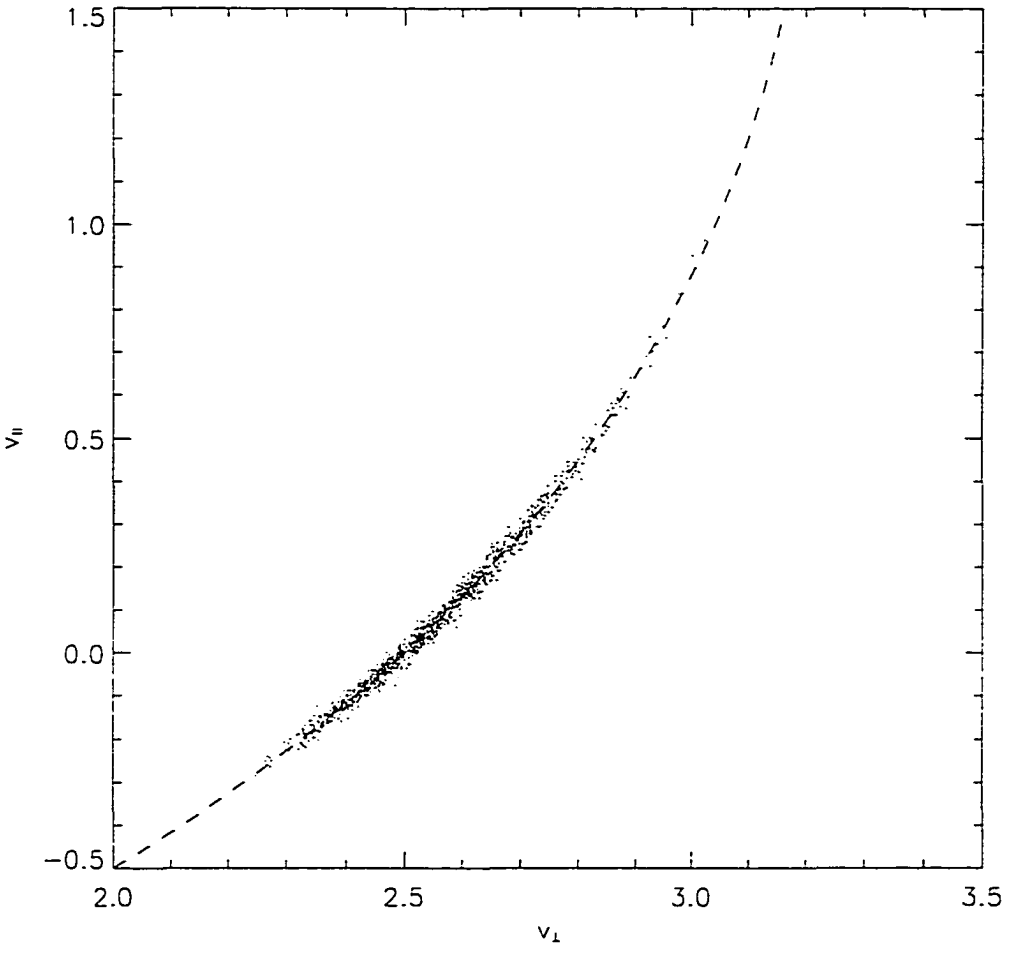


Figure 4.10: A phase space plot showing the results of a test particle scattering code. The dashed line shows the characteristic direction for the waves used in the simulation.

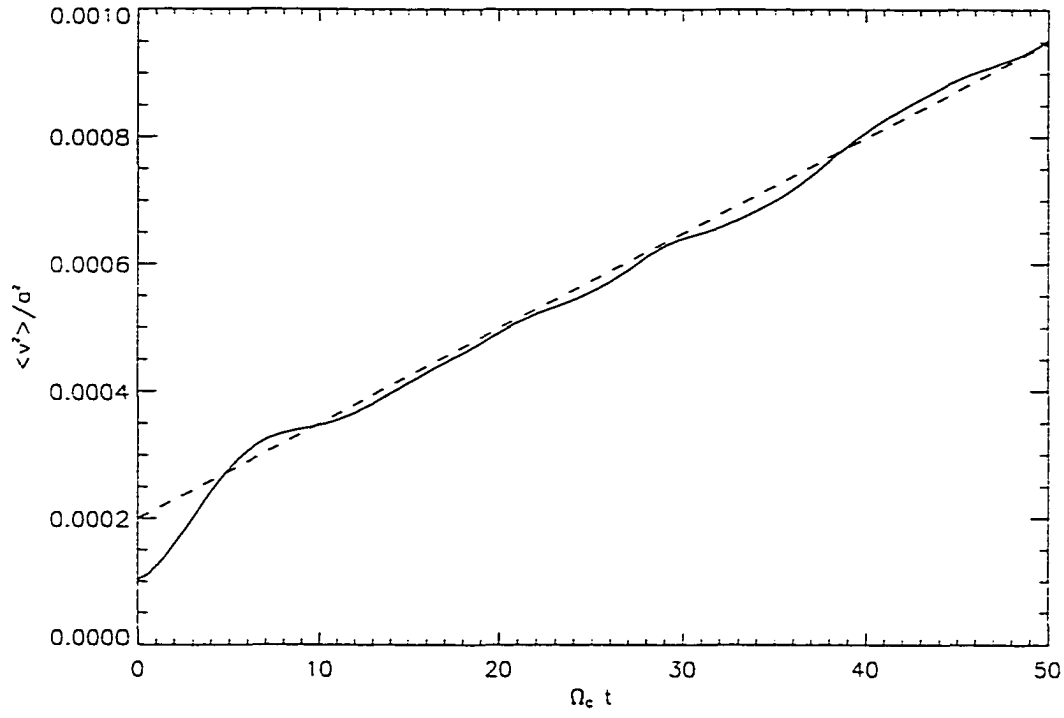


Figure 4.11: The rms parallel velocity of the test particles as a function of time. The dashes line is the velocity predicted by quasi-linear diffusion, with a 0.0001 a^2 offset.

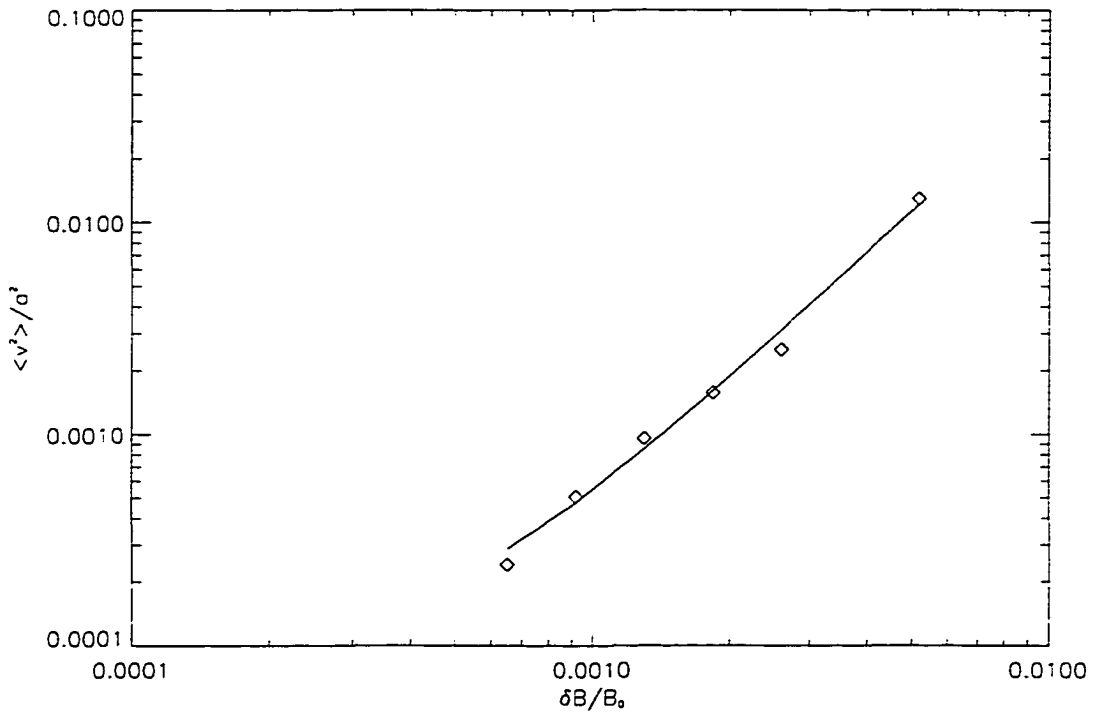


Figure 4.12: The rms parallel velocity of the test particles after 50 Ω_c^{-1} seconds, for several wave amplitudes. The line is the velocity predicted by quasi-linear diffusion.

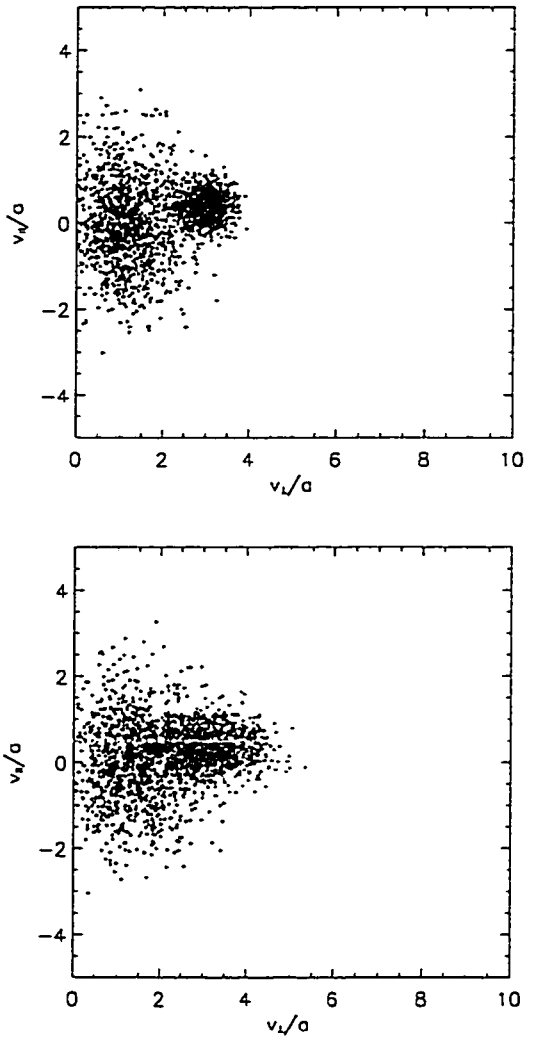


Figure 4.13: Velocity space distribution of the S^{+} test particles at t = 0 (top) and t = $10\Omega_c^{-1}$. The particles which were originally in the Maxwellian core are shown in blue and those originally in the ring-beam distribution are shown in red.

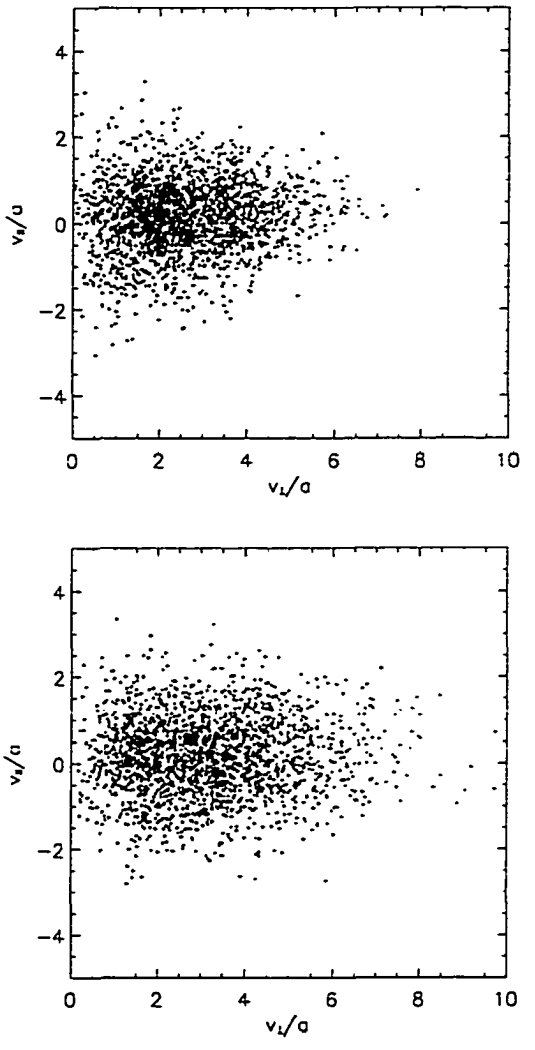


Figure 4.14: As figure 4.13, but for t = $25\Omega_c^{-1}$ (top) and t = $50\Omega_c^{-1}$ (bottom).

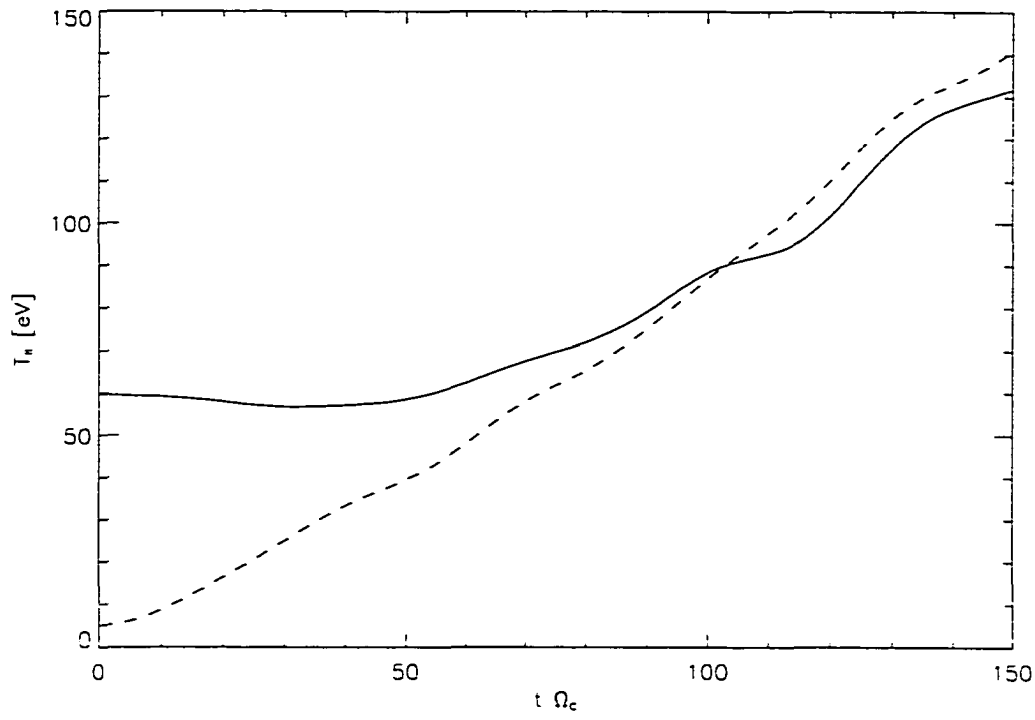


Figure 4.15: Parallel temperature of the test particles, for particles which were originally part of the Maxwellian population (solid line) and part of the ring-beam population (dashed line).

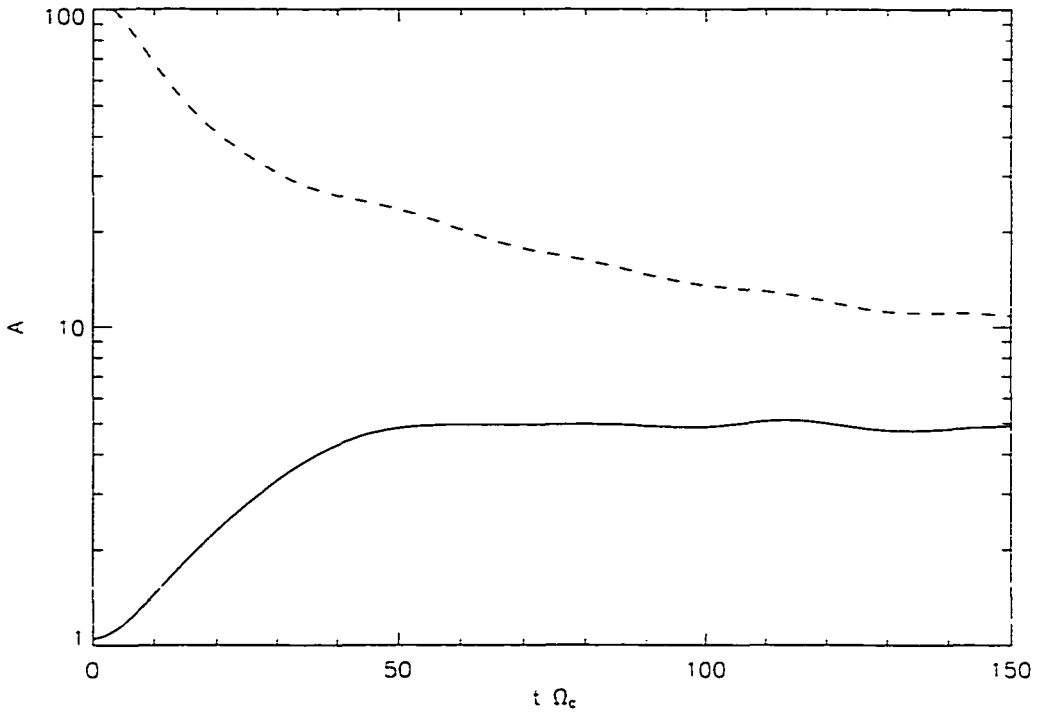


Figure 4.16: Thermal anisotropy of the test particles, for particles which were originally part of the Maxwellian population (solid line) and part of the ring-beam population (dashed line).

of particles remains unchanged for 500 Ω_c^{-1}. The oxygen gyrofrequency is sufficiently different from the wave frequencies, and the oxygen ions are not affected by non-resonant diffusion. This allows me to place an upper limit on the O^+ abundance. I solve equation 4.2 numerically for a background plasma plus oxygen ions in a ring-beam distribution of various parallel temperatures and densities, and $T_\perp = E_{pu} = 271$ eV. Since the oxygen ions are not scattered by the waves, they would retain their pick-up energy and whatever parallel temperature they had as neutrals. I vary density to find the maximum pick-up density, as a function of temperature, which would be stable. Figure 4.17 shows this upper limit on pick-up O^+ density.

The neutral oxygen atoms could have a relatively low temperature, $T \ll 1$ eV, since temperatures at the exobase are estimated to be under 1000 K [Strobel et al., 1994; Ballester et al., 1994; Kliore et al., 1975]. If the exosphere is the result of atmospheric sputtering, then the neutrals may attain a temperature less than or of order the energy with which they are ejected. Models of neutral loss and analysis of extended neutral clouds imply sputtered particles have a velocity of roughly 2-5 km/s [Smyth, 1992], which implies an ejection energy of 0.3 to 2 eV. If I adopt a value of 2 eV for the temperature, then the pick-up density would have to be under 212 cm^{-3} at all points along the Galileo trajectory. Adopting a lower temperature would result in a lower upper limit. Assuming a symmetric exosphere and a density of 212 cm^{-3} extending out to 6 R_{Io}, the maximum flux would be $2\pi(6R_{Io})^2 v_{cr} n_{max}$. This limit implies a source of less than 9×10^{27} oxygen ions per second from Io's exosphere. The actual flux should be much lower, since the oxygen density would not be uniform. If the column density of neutral oxygen was proportional to b^{-2}, then the upper limit on the source between 1.6 and 6 R_{Io} would be 4×10^{26} oxygen ions per second. Huddleston et al., 1998, estimated an exospheric source of 8×10^{26} SO_2^+ ions per second. The estimated source of ions supplying the Io torus as a whole is believed to be greater than 3×10^{28} ions per second (assuming a 2:1 oxygen:sulfur source rate) [Brown and Ip, 1981; Shemansky and Sandel, 1982; Smith and Strobel, 1985], much larger than my upper limit on the source from from Io's exosphere. I conclude that ionizations in the exosphere are a minor process, and that most of the plasma supplying the torus is either ionized closer to Io ($r < 1.6$ R_{Io}) [Bagenal, 1997] or well away from Io, in an extended neutral cloud [Smyth, 1992].

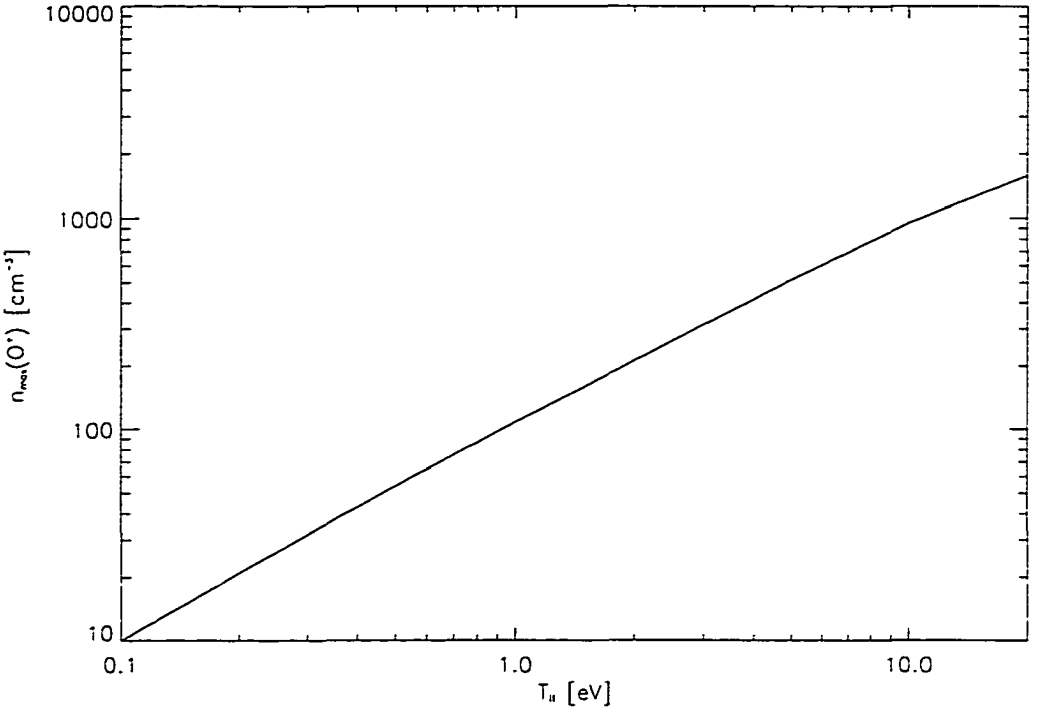


Figure 4.17: The upper limit on O^+ pick-up ion density, based on marginal stability of ion cyclotron waves. The background, O^+ density is 2000 cm^{-3}.

4.5 Io's neutral exosphere

Given these limits on the pick-up ion density, as well as the estimate of SO_2^+ pick-up density, I will now estimate the column density of Io's neutral exosphere, and compare the results to previous models and observations. The local density of pick-up ions is related to the source rate by integrating the source rate along a stream line

$$n_{\alpha+} = \int dt \frac{\partial n_{\alpha+}}{\partial t} = \int \frac{dy}{v_{flow}} \frac{\partial n_{\alpha+}}{\partial t} \tag{4.27}$$

$$\frac{\partial n_{\alpha+}}{\partial t} = \left(I_\alpha n_e + \sum_{\beta+N} \sigma_{(\beta+N,\alpha)}\, n_{\beta+N} < v > \right) n_\alpha \tag{4.28}$$

where I_α is the electron impact ionization rate coefficient, and $\sigma_{(\beta+N,\alpha)}$ the cross section for the charge exchange reaction, $\alpha + \beta^{+N} \rightarrow \alpha^+ + \beta^{+(N-1)}$ (e.g. $\sigma_{S^+,O}$ refers to the reaction $S^+ + O \rightarrow S + O^+$), and $< v >$ the average, relative velocity between α and β^{+N}. In the case of the waves seen on the inbound segment of Galileo's trajectory, the stream lines did not pass too close to Io and because the thermal ion velocity is small compared to the flow velocity, $< v > \sim v_{cr}$. Combining these equations, and assuming that the background plasma remains constant,

$$n_{\alpha+} = \int_{y_{sc}}^{\infty} dy \left(\frac{I_\alpha n_e}{v_{flow}} + \sum_{\beta+N} \sigma_{(\beta+N,\alpha)}\, n_{\beta+N} \right) n_\alpha \tag{4.29}$$

where where y is the direction parallel to the flow velocity and y_{sc} is position of the spacecraft. This may be rewritten in terms of the neutral column density, N_α:

$$\begin{aligned} n_{\alpha+} &= \left(\frac{I_\alpha n_e}{v_{flow}} + \sum_{\beta+N} \sigma_{(\beta+N,\alpha)}\, n_{\beta+N} \right) \int_{y_{sc}}^{\infty} dy\, n_\alpha \\ &= \left(\frac{I_\alpha n_e}{v_{flow}} + \sum_{\beta+N} \sigma_{(\beta+N,\alpha)}\, n_{\beta+N} \right) N_\alpha/2 \\ &\equiv N_\alpha/2L_\alpha \end{aligned} \tag{4.30}$$

$$N_\alpha = 2L_\alpha n_{\alpha+}$$

where L_α is a characteristic length for ionizations. L_α is related to the ionization rate by $L_\alpha = v_{flow}\tau_{ionization}$. If a flux tube advected through a uniform background of neutrals,

the pick-up ion density would equal the neutral density after the flux tube had traveled a distance L_α. N_α is defined to be the column density seen by an observer, $\int_{-\infty}^{\infty} n_\alpha dy$, and the factor of two comes from assuming Galileo's trajectory bisected the exosphere, which is reasonable $r \gg 1.4R_{Io}$. Figure 4.18 shows the geometry of the encounter and these column integrals.

Table 4.1 gives the cross sections that I will use [McGrath and Johnson, 1989]. The electron impact ionization rate is temperature dependent, and very sensitive to the super-thermal electrons (since the core electron temperature is less than the ionization potential) Schreier et al., 1998, give functional forms for I_S and I_O which assume that the super-thermal population can be approximated by a hot Maxwellian component. Similar data on I_{Na} can be found in Brown et al., 1983. The electron impact ionization cross section for SO_2 is 0.71 $Å^2$ for 20 eV electrons and 2.2 $Å^2$ for 90-125 eV electrons (personal communication, D. F. Strobel, 1997). Observational data on the super-thermal population is limited. As a result, the electron impact ionization rate is poorly constrained. For a 5 eV core temperature, hot component temperatures of 50 to 200 eV, and hot component abundances of 0.01 to 0.05, I find that

$$\begin{aligned} 2 \cdot 10^{-10} &< I_{SO_2} < 2.5 \cdot 10^{-10} cm^{-3} s^{-1} \\ 1.7 \cdot 10^{-9} &< I_O < 5.2 \cdot 10^{-9} cm^{-3} s^{-1} \\ 1.4 \cdot 10^{-9} &< I_S < 1.0 \cdot 10^{-8} cm^{-3} s^{-1} \\ 5.6 \cdot 10^{-8} &< I_{Na} < 6.1 \cdot 10^{-8} cm^{-3} s^{-1} \end{aligned} \tag{4.31}$$

The lower limits are essentially all due to the thermal electrons, so the range represents the uncertainty due to the poorly constrained superthermal population. The low SO_2^+ rate is a result of the fragile molecule: Electron impacts are more likely to ionize and dissociate the molecule (e.g. into O^+ and SO) than ionize it intact.

Combining these rates and cross sections with the plasma composition and density observed by Galileo [Crary et al., 1998; Frank et al., 1996; Gurnett et al., 1996], listed in Table 1.6, I obtain

$$\begin{aligned} 1.2 \cdot 10^{11} &< L_O < 1.6 \cdot 10^{11} cm \\ 9 \cdot 10^{10} &< L_S < 1.9 \cdot 10^{11} cm \\ L_{SO_2} &= 2.9 \cdot 10^{11} cm \\ 2.2 \cdot 10^{10} &< L_{Na} < 2.4 \cdot 10^{10} cm \end{aligned} \tag{4.32}$$

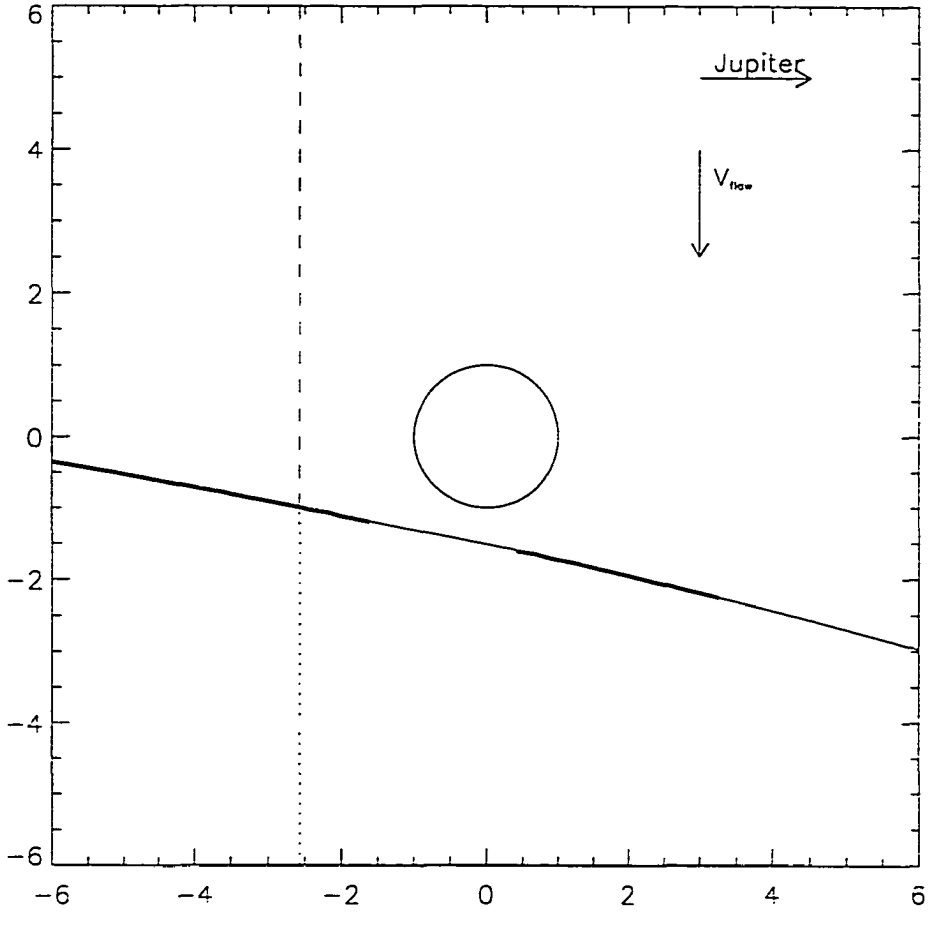


Figure 4.18: The geometry of the Galileo JO encounter with Io. The solid line shows the spacecraft's trajectory, with the thick line showing the regions in which ion cyclotron waves were observed. The dashed line shows the path integral used in equation 4.30, an approximate stream line. The dashed and dotted lines together show the line of sight for a terrestrial observer.

Reaction	Cross section $[cm^2 x 10^{16}]$
$O^+ + O \rightarrow O + O^+$	22
$O^{++} + O \rightarrow O^+ + O^+$	9.86
$S^+ + O \rightarrow S + O^+$	0.1
$S^{++} + O \rightarrow S^+ + O^+$	3.8
$O^+ + S \rightarrow O + S^+$	5.2
$O^{++} + S \rightarrow S^+ + O^+$	39
$S^+ + S \rightarrow S + S^+$	40
$S^{++} + S \rightarrow S^+ + S^+$	13.5
$O^+ + SO_2 \rightarrow O + SO_2^+$	11
$O^{++} + SO_2 \rightarrow O^+ + SO_2^+$	≈ 13
$S^+ + SO_2 \rightarrow S + SO_2^+$	<0.1
$S^{++} + SO_2 \rightarrow S^+ + SO_2^+$	≈ 16
$O^+ + Na \rightarrow O + Na^+$	0
$O^{++} + Na \rightarrow O^+ + Na^+$	49.2
$S^+ + Na \rightarrow S + Na^+$	0
$S^{++} + Na \rightarrow S^+ + Na^+$	57.2

Table 4.1: Charge exchange rates for Io's exosphere, from McGrath and Johnson, 1989.

The uncertainty in the L_{SO_2} due to the uncertain electron temperature, is only 3x10^9 cm, since the electron impact ionization rate is very low. For sulfur, 40-80% of the ionizations are from charge exchange, while 60-80% of the oxygen ionizations and 9-10% of the sodium ionizations are. Note that all of these lengths are a few hundred Io radii. This means that only a small fraction of Io's exosphere is lost on an advection time ($\sim 12\ R_{Io}/v_{cr} = 385$ sec.)

The density estimates of of Huddleston et al., 1997, imply a neutral SO_2 column density of 4.5x10^{13} cm^{-3} at an impact parameter of 1.6 R_{Io} in the anti-jovian direction, and decreases approximately as $b^{-3.1}$ I will not consider the outbound (sub-jovian) data: During that period, the upstream flow passed close to Io, and equation 4.28 assumes unperturbed flow.

No direct observations of Io's SO_2 exosphere exist. The lower atmosphere, with SO_2 column densities of order 10^{16} cm^{-2} has been observed [Ballester et al., 1994], but data on the exosphere is indirect. Schneider, 1988, observed the sodium exosphere and found the column density of sodium to be

$$N_{Na} = 2.6 \cdot 10^{12} \left(\frac{b}{R_{Io}}\right)^{-2.48} \text{cm}^{-2} \tag{4.33}$$

He then discussed the exosphere as a whole, by assuming that Na represented 1% of the exospheric density. Scherb and Smyth, 1993, reported observations of the 6300 Å line of neutral oxygen near Io, with roughly 9 R_{Io} resolution. They attributed they attributed the emissions to molecular dissociations,

$$\begin{aligned} SO_2 + e^- &\rightarrow SO + O + e^- \\ SO + e^- &\rightarrow S + O(^1D) \end{aligned} \tag{4.34}$$

since this process is more efficient at producing excited oxygen under plasma torus conditions. They found that a model of SO_2 outflow matched their observations as well as other observations of an extended neutral cloud. In this model, 8.9x10^{27} SO_2 molecules/s are ejected isotropically from Io's exobase, with a radial velocity of 2.6 km/s. The molecules were given dissociation life times of 1.4 hours (SO_2) and 1.0 hours (SO). While they did not present the resulting exospheric densities, this is relatively easy to calculate. Although they used a Monte Carlo model, which included many higher order effects, a first order approxi-

mation for distances $r < 5.8R_{Io}$ need only include Io's gravity and dissociation. The radial velocity of the neutrals is $v_r = \sqrt{v_0^2 - v_{esc}^2(R_{exo}/r)}$, where R_{exo} is the radius of, and v_{esc} the escape velocity at, the exobase. From this and conservation of flux, I calculate the model neutral column densities. Figure 4.19 shows the SO_2^+ density estimated by Huddleston et al., 1997, and the predicted values for the exosphere discussed by Schneider, 1988, and Scherb and Smyth, 1993. Scaling from the sodium column density gives reasonable agreement, for Na:SO_2 = 50 to 100 (reasonable close to the 1% value assumed by Schneider, 1988). However, the SO_2 column density decreases more sharply than the Na column density. This is consistent with an outflow and dissociation of SO_2 as the molecules move away from Io. On the other hand, the outflow model of Scherb and Smyth, 1993, gives too weak of a dependence on impact parameter. When I use dissociation times of 0.5 hours, for both SO_2 and SO, the slope of the column density is in reasonable agreement with data. However, the column density implied by the ion cyclotron waves is a factor of two to three higher than predicted by the outflow model. This could be due to preferential loss in the anti-jovian direction and total source rate similar to the value suggested by Scherb and Smyth. Time, longitude and local time variability could also account for the difference in magnitude: Scherb and Smyth reported a factor of two to three variability in exospheric [OI] emissions, as a function of local time and sub-jovian longitude.

Converting the neutral sodium density of Schneider, 1988, to a pick-up ion density, I obtain a peak density of 18 cm^{-3} pick-up sodium ions. This is sufficient to generate ion cyclotron waves with an amplitude of order 40 nT, which would have been easily detected. The absence of such waves indicates that thermalized sodium ions are present in the Io torus, however it is not possible to constrain this population. The uncertainties in thermal population temperature, and the exact distribution of the Na pick-up ions, are too large to allow meaningful constraints. Similarly, converting the neutral densities of the outflow model, using 0.5 hour lifetimes, predicts a density of 180 cm^{-3} pick-up oxygen ions or less, at all points along the Galileo spacecraft's trajectory. This is consistent with my upper limits on pick-up oxygen density.

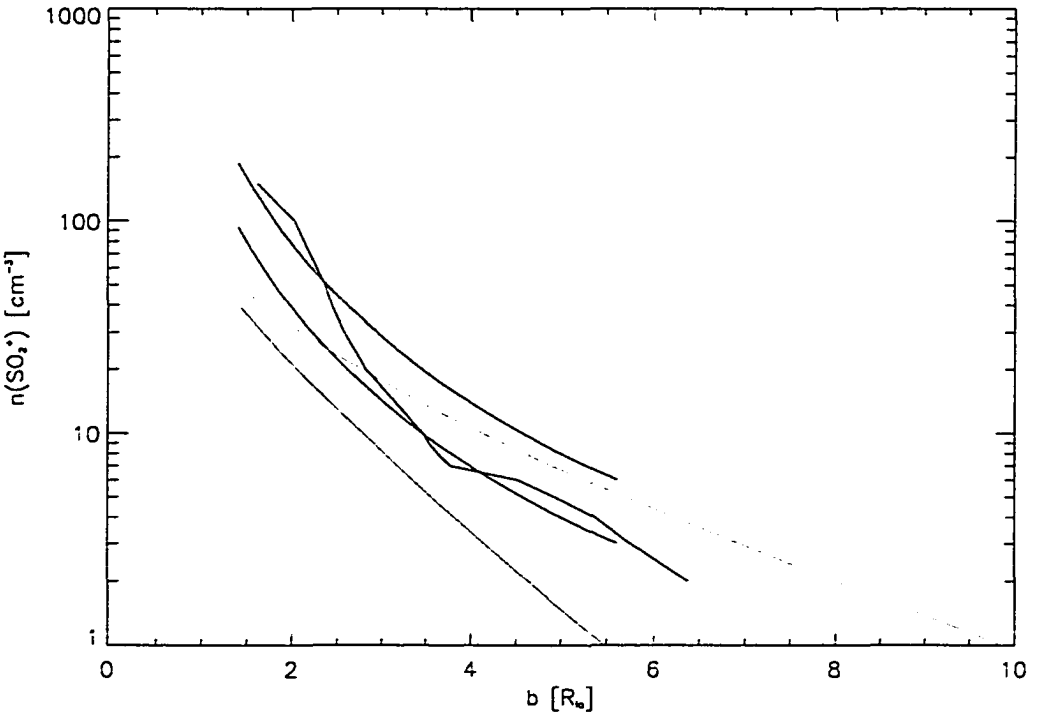


Figure 4.19: The SO_2^+ density profile of Huddleston et al., 1997, black. Colored lines show the predictions of various models. Extrapolating from the neutral Na density [Schneider, 1988] gives the purple ($Na:SO_2$ = 1:100) and green ($Na:SO_2$ = 50). The outflow model of Scherb and Smyth, 1993, results in the yellow (1.4 and 1 hour dissociation times) and orange (0.5 hour dissociation times).

4.6 Summary

In this chapter, I examined the strong ion cyclotron waves observed near Io by the Galileo spacecraft, and used them to estimate the exospheric source of ions and properties of Io's exosphere. These waves are generated by pick-up ions, which initially have an unstable, ring-beam distribution. While a pure ring-beam distribution is always unstable, the presence of a thermal background of ions causes damping and can prevent wave growth. Ion cyclotron waves were observed near the SO_2^+ gyro-frequency, but not the S^+ or O^+ gyro-frequencies. This is a result of the short lifetime of the molecular ion: SO_2^+ dissociated on a time scale of order an hour, and does not survive long enough to evolve into a thermal distribution. As a result, pick-up SO_2^+ ions can easily generate ion cyclotron waves. In contrast, the S^+ and O^+ ions of the Io plasma torus are a dense, thermalized population which prevents the S^+ and O^+ pick-up ions from generating ion cyclotron waves.

I discussed the basis for and limits of quasi-linear diffusion, and showed that this technique cannot be used to model the evolution of pick-up SO_2^+ ions. Quasi-linear diffusion assumes that the wave amplitudes are sufficiently small, and that the Vlasov equation may be separated into two equations. (One describing an oscillation at the wave frequency and one describing a much slower, diffusive evolution of the distribution function.) I showed that the observed wave amplitudes are too strong for this approximation to be justified. In addition, I examined the role of non-resonant diffusion and estimated that S^+ ions would be subject to significant, non-resonant scattering. Given the observed amplitudes, S^+ is affected by waves with roughly 1 Hz of the resonant frequency. I then considered the scattering of particles along characteristic directions. So long as $\vec{k} \cdot \vec{E} = \vec{k} \times \vec{B}_0 = 0$, the ions scatter along the line, $v_\perp^2 + (v_\parallel - \omega/k)^2 =$ constant. This does not depend on the amplitude of the waves, nor any of the assumptions that go into quasi-linear theory, but does require a non-dispersive wave spectrum. In the case of pick-up ions, the instability would produce waves with $\vec{k}$ parallel and anti-parallel to the background, magnetic field. I showed that this bi-directional spectrum alters the character of particle scattering. Rather than scattering along a characteristic, the particles in a ring-beam distribution will diffuse in both energy and pitch angle. I find that the Alfvén Mach number is a rough guide to the character of the scattering. When the Alfvén Mach number is large, the particles rapidly scatter in

pitch angle, but their energy change slowly. This is the case for cometary pick-up ions and is reasonably well approximated by scattering along characteristic directions and the formation of bi-spherical distributions. However, when the Alfvén Mach number is of order unity, or smaller, the particles experience anisotropic diffusion and evolve in both energy and pitch angle. The relative rate of the parallel versus perpendicular diffusion is equal to v_{flow}^2/v_o^2, which is roughly the square of the Alfvén Mach number. As a result, pick-up ions near Io should evolve into an anisotropic "pancake" distribution, rather than a shell or bi-spheric distribution, and have an anisotropy of order $1/M_A^2$. The same process may be the cause of highly anisotropic ions observed in the magnetosphere of Saturn.

The absence of S^+ and O^+ waves shows that the thermal core population is able to suppress the instability caused by pick-up ions. This places constraints on the relative densities and temperatures of the core and pick-up ions. I discussed various ways to estimate these limits, using the dispersion relation to calculate marginal stability criteria. I find that these limits are are very sensitive to the parallel temperature of the pick-up ions. To constrain this value, I performed test particle integrations. As predicted by my earlier estimates, the S^+ test particles experienced strong, non-resonant scattering and evolved into an anisotropic distribution rather than a bispheric shell. The core S^+ ions also scattered into an anisotropic distribution and were strongly heated by the waves. Since I used the observed spectrum, and did not include wave growth and damping by the ions, these simulations are only accurate for roughly the first 10 Ω_c^{-1}, or 2 seconds. At that time, the core population had absorbed 10% of the wave energy, and my assumption of a constant wave spectrum becomes invalid. In contrast, the oxygen test particles were not affected by the waves at all. Their gyro-frequency is sufficiently different from the wave frequency, and no non-resonant scattering occurs. As a result, the O^+ pick-up ions would retain their initial temperature and anisotropy. I estimate that this initial temperature would be under 2 eV. Combining this with the pick-up velocity, and using the ion cyclotron wave dispersion to calculate a marginal stability criteria, I find that the density of O^+ pick-up ions must be under 200 cm^{-3} at all points along the Galileo trajectory. This corresponds to a source of less than $5x10^{27}$ oxygen ions per second. Since this is under 25% of the total oxygen source which produces the Io torus, I conclude that Io's exosphere is a minor source for the torus. Most of the ionizations either occur closer to ($r < 1.6\ R_{Io}$) or farther from ($r > 10\ R_{Io}$) Io.

Given this limit on O^+ pick-up densities, and the SO_2^+ pick-up density inferred by Huddleston et al. (1997) from the ion cyclotron wave amplitude, I estimate the density of Io's neutral exosphere. I relate the pick-up ion density to the source rate, integrated along a stream line. Since the source rate is proportional to the neutral density, this allows me to convert the pick-up ion density to the integrated column density of neutrals. Using this technique, I find that the neutral oxygen column density is under $3x10^{13}$ cm^{-2}. The neutral SO_2 density is approximately $1.9x10^{14}$ $(b/R_{Io})^{-3.1}$. This is in agreement with estimates based on sodium observations [Schneider, 1988], assuming a sodium to sulfur dioxide ratio of 1:50 at the exobase and decreasing to 1:100 at 5 R_{Io}. This ratio may be lower than expected, since it implies a sodium abundance of under 1%, once SO_2 has dissociated into S and O. It also matches the model of sulfur dioxide outflow and dissociation [Scherb and Smyth, 1993], if I assume a dissociation life time for SO and SO_2 of 0.5 hours rather than 1.0 and 1.4 hours.

Chapter 5

Summary

5.1 Results

In this thesis, I have modeled Io's interaction with the jovian magnetosphere and developed models which explain how the interaction results in electron acceleration. The results are consistent with observations, and produce energetic electrons with roughly the correct particle energy and power to account for Io's auroral features and Jupiter's decametric radio emissions. I have also used the Galileo observations of ion cyclotron waves near Io to estimate the properties of Io's exosphere.

In chapter 2, I developed a model of Alfvén wave propagation and particle acceleration due to non-ideal MHD effects. When electron inertia is included, Alfvén waves have a parallel electric field when $k_\perp c/\omega_{pe}$ is of order one or larger, a condition which occurs in the low density region between the Io torus and Jupiter's ionosphere. When I apply this model to the initial Alfvénic disturbance generated by Io, I find that the Alfvénic disturbance would be almost completely reflected by the non-uniform density and magnetic field strength along Io's flux tube. In the process, however, the disturbance produces electron beams, due to the parallel electric field and Fermi acceleration. I estimate that these electrons have a broad spectrum, with energies of a few to a few hundred keV. The total power carried by these electrons is approximately 10^{11} W. These values are consistent with the precipitating particles required to produce Io's auroral spot in the jovian atmosphere. I also find that a small fraction of the accelerated electrons mirror, and travel back up the field line in a loss cone distribution with an average energy of 1-10 keV. This matches recent observations of S-burst decametric emissions, which infer a source of few keV electrons traveling away from Jupiter in a loss cone distribution. When I assume that the electrons have a Maxwellian distribution, my model predicts that less than $6\text{x}10^9$ W goes into these electrons. The saturation efficiency of the cyclotron maser instability, the probable emission mechanism, is only a few percent, and the power I infer is insufficient to account for the S-bursts. In contrast, when I assume the electrons are in a κ distribution, i.e. approximately a Maxwellian with a high energy tail, a significantly larger fraction of the electrons mirror, and the power going into mirroring electrons increases to $2\text{x}10^{10}$ W, which is sufficient to power the S-bursts assuming a conversion efficiency of 2.5%. The cyclotron maser instability has a saturation efficiency of a few percent [Wagner et al., 1984; Genova et al., 1989, and references therein].

In chapter 3, I examined the acceleration of particles by a steady potential structure, which would be present on the downstream side of Io and in the wake. After considering and dismissing particle acceleration in the collisional ionosphere of Jupiter, I modeled the role of double layers. I assumed that magnetospheric double layers can not alter their environment, due to the large scales of the system relative to the double layer, and that the flux of particles approaching the double layer is a fixed boundary condition. Under this assumption I use an analytic model of double layers to show that a single double layer can bifurcate when the potential across it becomes greater than a few times the electron temperature. I conclude that multiple double layers are likely to form along a given field line, assuming the criteria for the Buneman instability is satisfied, but that strong double layers, with $\Phi \gg T_e/e$ are unlikely to occur in a magnetosphere. Applying this theory to Io and Io's wake, I showed that a significant length of the field lines would be unstable and become populated with double layers, if the parallel current density exceeded roughly 10^{-7} A/m^2. I adopted an approximate model in which double layers act to limit this current and keep it at or below a threshold value, which I took to be 10^{-7} A/m^2. As a result, the steady current closing through Io is limited to roughly $4 \cdot 10^4$ A, which is orders of magnitude weaker than the Alfvénic signature observed by Voyager. Despite dissipating a trivial amount of power, the double layers would act to accelerate electrons, and the energy and flux predicted by my model is in reasonable agreement with the bi-directional electron beams observed by Galileo, immediately downstream of Io. Next, I applied my model of double layers to Io's wake. Re-accelerating this nearly-stagnated flow to co-rotation velocity requires a substantial current. If no double layers formed, this would cause 5×10^{12} W of Joule heating in Jupiter's ionosphere. Allowing for the formation of double layers, I showed that approximately 10^{12} W of this power would go into accelerating electrons, instead of heating the ionosphere. The presence of double layers also increases the length of Io's wake, by limiting current flow and therefore the JxB force re-accelerating the wake. These results are consistent with the arc of auroral emission, which extends at least 90° downstream of the foot of Io's flux tube. If a few percent of these electrons mirrored rather than precipitating into Jupiter's atmosphere, they would represent enough power to generate the bulk of Io's decametric radio emissions.

In chapter 4, I considered the strong ion cyclotron waves, observed near Io by the Galileo spacecraft. These waves are caused by the unstable velocity distribution of recently

ionized pick-up ions. I showed that quasi-linear theory can not be used to model the evolution of SO_2^+ or S^+ pickup ions, due to the large amplitude of the waves. Instead, I numerically integrated the trajectories of S^+ and O^+ ions, to estimate the degree to which these ions are non-resonantly scattered by the observed waves. I found that both the thermal core and pick-up S^+ ions are strongly scattered, on time scales of order 10 $\Omega_{c,S+}^{-1} \sim 2$ sec. The pick-up ions scatter into a more isotropic distribution, which the core becomes moderately isotropic. Both components are strongly heated. This implies that the S^+ population causes significant wave damping, and modifies the spectrum of the waves. In contrast, O^+ ions are not scattered at all: Their gyro-frequency is too different from the wave frequencies. I use this fact, and the absence of ion cyclotron waves near the O^+ gyro-frequency, to place limits on the abundance of O^+ pick-up ions. The absence of waves implies that wave growth is stabilized by the background, thermal population of O^+. I calculate a marginal stability criteria using the ion cyclotron dispersion relation, and show that the density of O^+ pick-up ions was less than 200 cm^{-3} along the Galileo trajectory. I then estimate the properties of Io's neutral exosphere using this limit and the SO_2^+ density inferred by Huddleston et al. (1997) from the ion cyclotron wave amplitude. I relate the neutral column density to the local pick-up ion density and the ionization rate. From estimates of the ionization rates, I conclude that the column density of O is under $3x10^{13}$ cm^{-2} and the SO_2^+ column density is approximately $1.9x10^{14}(b/R_{Io})^{-3.1}$ cm^{-2}, above 1.6 R_{Io}. I compared this to previous estimates of Io's exosphere. The SO_2^+ is in reasonable agreement with Schneider (1988), where the total exospheric density was estimated from the observed Na density and an assumed mixing ratio of 1%. My estimates, based on the ion cyclotron waves and the Schneider Na profile, imply a SO_2:Na ratio of 50 at 1.6 R_{Io} increasing to 100 at 6 R_{Io}. When I assume an SO_2 and SO lifetime of 0.5 hours. my estimated column densities also agree with the results of Scherb and Smyth, 1993, who used OI emissions to constrain the exosphere.

5.2 Future modeling

My models are by no means a complete or comprehensive theory of the phenomena associated with Io's magnetospheric interaction. Further theoretical analysis and observa-

tions are necessary to improve the accuracy of the models. Nor have I addressed the small scale structure and rapid time variability, seen in auroral and radio emissions (e.g. the shape and spacing of decametric arcs.) I plan to continue work on theoretical models to address these issues.

The analytic model of double layers, which I applied in chapter 3, does not describe all of the important processes. The particles accelerated by the double layer would be subject to beam instabilities, causing wave turbulence as observed in laboratory experiments. To make quantitative comparisons to the electron beams observed in Io's wake by Galileo, and to estimate the fraction of these particles outside the loss cone (and, therefore a source of decametric emission), I plan to apply electrostatic particle in cell (PIC) simulations. These simulations will also provide insight into the formation and spacing of multiple double layers. The need to simulate thousands of Debye lengths to model multiple double layers may limit some of these simulations to one dimension. I have been discussing collaborations with Meers Oppenheim, and the problem appears to be computationally tractable.

In my analysis of the ion cyclotron waves near Io, I found that quasi-linear diffusion and test particle integrations can not accurately describe the evolution of SO_2^+ and S^+ ions. The amplitude of the observed waves is too large for quasi-linear models of SO_2^+ evolution, and S^+ ions experience strong non-resonant scattering which would affect the waves' amplitude and spectrum. These issues affect estimates of the SO_2^+ pick-up ion abundance and prevented me from placing limits on the S^+ pick-up abundance. I plan to address this problem using electromagnetic PIC simulations, which have previously been used to model the evolution of pick-up ions near comets [Gray et al., 1996]. P. C. Gray and I have discussed this matter, and we plan to begin work on it shortly. In addition, these simulations will allow modeling of the mirror mode waves observed closer to Io, and predictions for the Galileo encounters with Io in late 1999. As discussed below, I anticipate that lower hybrid waves, as well as ion cyclotron waves, may be observed during the I25 encounter in November, 1999.

I will also use MHD simulations of the Io interaction to study the effects of wave reflection within the Io torus and the evolution of the interaction from a transient Alfvénic disturbance to a steady current loop. Previous MHD simulations have assumed that Io moves through a uniform, infinite plasma. As I showed in chapter 2, the non-uniform density of the torus and the increasing magnetic field strength as the waves propagate away from Io

cause a large fraction of the wave energy to be reflected. The effect of these waves on the Io interaction is unknown, but could result in an enhancement of the magnetic field signature near Io. Unexpectedly strong magnetic field perturbations near Io were observed by the Galileo spacecraft, and attributed to either an intrinsic magnetic field or an extremely strong source of plasma. Wave energy trapped within the torus and along the Io/wake flux tubes may be an alternative explanation. In addition, the evolution of the Io interaction, from an Alfvénic structure with currents closing within the plasma, to a steady current loop closing within Jupiter's ionosphere, is poorly understood and not analytically tractable. In collaboration with Fran Bagenal and Jon Linker, I will be adapting the MHD models which Linker has previously used to study the Io interaction, to include the non-uniform plasma and the ionosphere of Jupiter.

5.3 Future observations

In recent years, observations of the jovian magnetosphere and the Io interaction have greatly improved, and allowed me to develop models of the acceleration process, as well as estimating the properties of Io's exosphere. As new observations become available, I plan to refine and improve (or discard and replace, if appropriate) my theories of the Io interaction. In this section, I summarize the observations which will, or could, be made in the next few years.

Radio emissions are now able to resolve the fine structure of S-burst emissions, and determine the parallel velocity of the source electrons [Zarka et al., 1996; Carr et al., 1997], and work on understanding this population continues. Recently, data from the WIND spacecraft has been used to extend the frequency range of observations down to 1 MHz, and estimate the emission cone angle and source location of the L-burst emissions [Queinnec and Zarka, 1998]. A determination of the decametric source location will allow better models of the emission process, source particles and the association between the radio emissions and aurora.

Observations of the jovian aurora also provide progressively better constraints on the Io interaction. The Hubble Space Telescope's Space Telescope Imaging Spectrometer (STIS) has taken images of both the Io spot and extended arc, with far higher resolution

and sensitivity than previous observations. However, only four images have been obtained to date, in contrast to the 45 images from Hubble's WFPC2 instrument in 1994-5 [Clarke et al., 1996] including a set of 11 taken over a six hour period, which provided data on time evolution of the aurora [Ballester et al., 1996]. The earlier WFPC2 coverage shows what may be available in the future from STIS, i.e. numerous images providing significant time resolution (e.g. 300 seconds and less). In addition, images of Io's auroral features were taken by the Galileo spacecraft, both in the visible and IR, and analysis of these data is in progress. The IR observations include spectra from 0.7 to 5 μm with 0.01 μm resolution. In addition to improved observations, comparisons between IR, visible and UV brightness have not yet been made, and would provide information on the energy and flux of precipitating particles.

The Cassini mission offers a very rich opportunity to study the jovian aurora and the Io interaction remotely. In December, 2000, Cassini will make a distant encounter with Jupiter. At this time, it would be possible to acquire high resolution images of the jovian aurora, in the UV, visible and IR wavelengths. The UV images would be of particular interest, since the Galileo spacecraft's UV spectrometer does not provide resolved images and has difficulty separating the Io features from the main auroral oval. In addition to imaging, the spacecraft's plasma wave spectrometer could observe the decametric radio emissions up to a frequency of 16 MHz (the similar instrument on Galileo is limited to frequencies of under 5 MHz, preventing significant observations of the decametric emissions.) The analysis of Voyager radio observations has shown the importance of observing jovian radio emissions while near Jupiter, and at latitudes and local times that can not be observed from the Earth.

While Earth-based observations will provide information on Jupiter's radio emissions and aurora, spacecraft are required for in-situ measurements of plasma conditions and magnetic fields. The Galileo Europa Mission (despite its name) will greatly improve our understanding of the interaction near Io. Beginning in May, 1999, the spacecraft's perijove will be reduced over the course of four orbits, allowing two close encounters with Io in October and November of 1999. During the perijove reduction phase of the mission, the spacecraft will conduct a detailed survey of the Io torus. The Io encounters include a close (500 km altitude) upstream pass on the I24 orbit (October 11) and a pass over Io's south pole and through the Io flux tube on the I25 orbit (November 26). I expect the I24 encounter to be similar in many ways to the previous encounter in December 1995. Of course, the wake

signature observed in 1995 will not be encountered. However, the strong ion cyclotron waves should be present, as should accelerated flow along the flanks of Io and possibly the mirror mode waves. The wave amplitude would be lower, since the trajectory is roughly 2.5 R_{Io} farther upstream than the J0 trajectory. This means the advected flux tubes will have spent 80 second less time in the source region. The density of pick-up ions, and therefore the amplitude of ion cyclotron and mirror mode waves will be lower.

In contrast, I predict that the I25 encounter will be radically different. First, by passing directly over the south pole of Io, the spacecraft will measure the currents flowing along the Io flux tube and magnetometry will resolve the question of whether Io has an intrinsic magnetic field. The fronts of multiple, reflected Alfvén waves may also be observed, providing information on the evolution of the interaction as waves reflected off Jupiter's ionosphere return to Io. Second, the encounter occurs at a (jovian) longitude of 150°, where the field lines are more vertical (closer to perpendicular to the flow velocity) and Io is above the bulk of the Io torus. This will have at least two significant effects. The upstream conditions and the Alfvén wave propagation time to Jupiter will be very different from those of the previous encounters. This will alter the character of the interaction, as well as the relative abundance of pick-up ions. In addition, the vertical field geometry may result in lower hybrid, rather than ion cyclotron, waves. Pick-up ions may generate lower hybrid waves, but only if the beam velocity is small. Since the magnetic field will be more vertical during the I25 encounter, the pick-up ion's beam velocity will be lower than it was during the J0 encounter, and the ions may drive lower hybrid growth.

Although the Galileo encounters will greatly enhance our understanding of the interaction near Io, the plasma conditions at mid- and high latitudes remain poorly known. This represents the greatest uncertainty for models of the Io-Jupiter interaction. Almost no spacecraft data are available. The Voyager and Galileo observations are all from the low latitude regions of the jovian magnetosphere. Ulysses measured plasma conditions at high latitudes, but its trajectory never took it closer to Jupiter than roughly L = 7, well outside Io's L-shell. The only in situ data on the high latitudes of Io's L-shell comes from the Pioneer spacecraft. Pioneer 10 passed above and below the Io torus, and through Io's L-shell, at a latitude of approximately 20 degrees. This is still within the Io torus, rather than proton dominated plasma of the mid- and high latitudes, but provides interesting information on the

vertical structure of the Io torus. Pioneer 11 crossed Io's L-shell at a much higher latitude, roughly 50°, but I have been unable to find any published results on the plasma conditions near Io's L-shell. This may be a result of the 100 V threshold of the spacecraft's plasma analyzer instrument: Useful measurements of a relatively cool, low density H^+ plasma may not have been possible.

If studies of the Earth's magnetosphere are a guide, a spacecraft in a polar orbit will be necessary to fully understand the Io-Jupiter interaction, and the physics of the jovian aurora and radio emissions. Such a mission would also be of great value to studies of the jovian atmosphere as well as the Earth's magnetosphere, since the same physical processes are at work in a very different parameter regime. A jovian polar orbiter, with instruments along the lines of the POLAR spacecraft, would not be technically difficult. The propulsive requirements to enter a low, eccentric orbit around Jupiter are minimal, significantly less than orbital insertion for Galileo, due to the lower perijove, and immensely less than the requirements for the planned Europa Orbiter mission. A number of possible orbits allow a spacecraft to pass over the auroral region, within 2 jovian radii of the planet, without entering the dangerous radiation environment of the Io torus or inner radiation belts. These orbits allow measurements of Io's L-shell, as well as the acceleration processes in the Io and wake flux tubes, at multiple latitudes, as well as passing through the L-shells of the main auroral oval. In addition to in situ measurements, such a spacecraft could monitor Jupiter's aurora and atmosphere at less than 1 kilometer resolution, possibly as much as 100 meter resolution depending on the spacecraft's orbit. In comparison, the best resolution provided by the Galileo prime mission is 13 km. Although such a mission would be more difficult than a similar mission to study the Earth's magnetosphere, it is, debatably, on par with an ambitious Discovery mission.

Io's interaction with the jovian magnetosphere has been studied for the past 35 years, when Io's control of decametric radio emissions was noted. Since then, the Voyager and Galileo spacecraft have provided information on the interaction near Io, and the plasma conditions in the Io torus. Recent observations of the jovian aurora, and Io's auroral features, have documented another high-latitude consequence of the interaction. As we obtain more information, theories and models of the interaction continually improve. At the same time, more and better observations also reveal more complex phenomena and processes than

previously expected, presenting a constant challenge to theories and models.

Bibliography

Acuña, M.H., F.M. Neubauer, and N.F. Ness, Standing Alfvén wave current system at Io: Voyager 1 observations, *J. Geophys. Res.*,*86*, 8513, 1981.

Acuña, M.H., K.W. Behannon, and J.E.P. Connereny, Jupiter's magnetic field and magnetosphere, in *Physics of the Jovian Magnetosphere*, edited by A. J. Dessler, Cambridge University Press, Cambridgem 1983.

Bagenal, F., Alfven wave propagation in the Io plasma torus, *J. Geophys. Res.*,*88*, 3013. 1983.

Bagenal, F., Empirical model of the Io plasma torus: Voyager measurements, *J. Geophys. Res.*,*99*, 11043, 1994.

Bagenal, F., The ionization source near Io from Galileo wake data, *Geophys. Res. Let.*,*24*, 2111, 1997.

Bagenal, F., and J.D. Sullivan, Direct plasma measurements in the Io torus and inner magnetosphere of Jupiter, *J. Geophys. Res.*,*86*, 8447, 1981.

Ballester, G. A., et al., Detection of the SO_2 atmosphere on Io with the Hubble Space Telescope, *Icarus. 111*. 2. 1994.

Ballester, G. A., et al., Time-resolved observations of Jupiter's far-ultraviolet aurora, *Science, 274*, 409, 1996.

Barbosa, D.D., et al., Voyager observations of lower hybrid noise in the Io plasma torus and anomalous plasma heating rates, *Ap. J., 289*, 392, 1985.

Barnett, A., In situ measurements of the plasma bulk velocity near the Io flux tube, *J. Geophys. Res.*,, 91, 3011, 1986.

Barnett, A., and S. Olbert, Radiation of plasma waves by a conducting body moving through a magnetized plasma, *J. Geophys. Res.*,*91*, 10117, 1986.

Baron, R., et al., Imaging Jupiter's aurorae from H_3^+ emissions in the 3-4 μm band, *Nature, 353*, 539, 1991.

Belcher, J.W., et al., Plasma observations of the Alfvén wave generated by Io, *J. Geophys. Res.*,*86*, 8508, 1981.

Bigg, E.K. Influence of the satellite Io on Jupiter's decametric emissions, *Nature, 203*, 1008, 1964.

Borovsky, J.E., The strong-double-layer model of auroral arcs: An assessment, in *Auroral Plasma Dynamics*, ed. R.L. Lysak, 1993.

Bostrom, R.G., et al., Characteristics of solitary waves and weak double layers in the magnetospheric plasma, *Phys. Rev. Lett., 61*, 82, 1988.

Broadfoot, A.L., et al., Overview of the Voyager ultraviolet spectrometry results through Jupiter encounter, *J. Geophys. Res.,86*, 8259, 1981.

Brown, R.A., and W.-H. Ip, Atomic clouds as distributed sources for the Io plasma torus, *Science, 213*, 1493, 1981.

Carr, T.D., M.D. Desch, and J.K. Alexander, Phenomenology of magnetospheric radio emissions, in *Physics of the Jovian Magnetosphere*, edited by A. J. Dessler, Cambridge University Press, Cambridgem 1983.

Carr, T.D., et al., Recent results by the University of Florida group from low frequency radio observations of Jupiter and Neptune, in *Planetary Radio Emissions IV*, Austrian Academy of Science Press, Vienna, 1997.

Clarke, J.T., et al., Far-UV imaging of Jupiter's aurora and the Io footprint with the HST WFPC 2", J.T. Clarke et al., *Science, 274*, 404, 1996.

Clarke, J. T., et al., HST observations of Jupiter's UV aurora during Galileo in situ measurements, presented at the American Geophysical Union's Fall Meeting, San Francisco, December 1997.

Connerney, J.E.P., Doing more with Jupiter's magnetic field, in *Planetary Radio Emissions III*, Rucker, Bauer and Kaiser, ed., 1991.

Connerney, J.E.P., New models of Jupiter's magnetic field constrained by the Io flux tube footprint, submitted to *J. Geophys. Res.*, 1998.

Connerney, J.E.P., et al., Images of excited H_3^+ at the foot of the Io flux tube in Jupiter's atmosphere, *Science, 262*, 1035, 1993.

Connerney, J.E.P., et al., H_3^+ emissions at the foot of the Io flux tube, paper presented at 27 Annual Meeting, Div. Planet. Sci., Mauna Lani, Hawaii, Oct. 9-13, 1995.

Crary, F.J., On the generation of an electron beam by Io, *J. Geophys. Res.,102*, 37, 1997.

Crary, F.J. and F. Bagenal, Coupling the plasma interaction at Io to Jupiter, *Geophys. Res. Let.,24*,2135, 1997.

Crary, F.J., et al., Anisotropy and proton density in the Io plasma torus derived from whistler wave dispersion, *J. Geophys. Res.,101*, 2699, 1996.

Crary, F.J., et al., Galileo plasma spectrometer measurements of composition and temperature in the Io plasma torus, submitted to *JGR*, Feb. 1998.

Dulk, G. A., Io-related radio emission from Jupiter, Ph.D. thesis, University of Colorado, 1965.

Erdélyi, A., et al., *Higher Transcendental Functions*, vol. 1, chap. 3, McGraw-Hill, New York, 1953.

Ferraro, V.C.A., and C. Plumpton, Hydromagnetic waves in a horizontally stratified atmosphere, *Astroph. J., 127*, 459, 1957.

Frank, L.A., et al., Plasma observations at Io with the Galileo spacecraft, *Science, 274*, 394, 1996.

Galeev, A.A., et al., The nonlinear theory of the Buneman instability, *Sov. Phys. JETP, 54*, 306, 1981.

Genova, Françoise, et al., Jupiter decametric radiation, in *Time Variable Phenomena in the Jovian System*, NASA SP-494, Belton et al., ed., 1989.

Goertz, C.K., Io's interaction with the plasma torus, *J. Geophys. Res.,85*, 2949, 1980.

Goertz, C.K., Kinetic Alfvén waves on auroral field lines, *Planet. Space Sci., 32*, 1387, 1984.

Goertz, C.K., and R.W. Boswell, Magnetosphere-ionosphere coupling, *J. Geophys. Res.,84*, 7239, 1979.

Goertz, C.K. and P.A. Deift, Io's interaction with the magnetosphere, *Planet. Space Sci., 21*, 1399, 1973.

Goldreich, P., and D. Lynden-Beil, Io, A Jovian unipolar inductor, *Astroph. J., 156*, 59, 1969.

Goldstein, M.L., et al., A theory of Jovian decametric radiation, *J. Geophys. Res.,88*, 792, 1983.

Goldstein, D.A., and C.K. Goertz, Theories of radio emissions, pp. 317-352 in *Physics of the Jovian Magnetosphere*, edited by A.J. Dessler, Cambridge University Press, Cambridge, 1983.

Gray, P. C., et al., Heating of the solar wind py pickup ion driven Alfvén ion cyclotron waves, *Geophys. Res. Let.,23*, 113, 1996.

Gurnett, D.A., et al., Galileo plasma wave observations in the Io plasma torus and near Io, *Science, 274*, 391, 1996.

Guyot, M., and Ch. Hollenstein, Experiments on potential gradients in a current-carrying plasma. I. Potential structures, *Phys. Fluids, 26*, 1596, 1983.

Hansen, C., et al., Dynamic and Debye shielding and anti-shielding, *Phys. Plasmas, 3*, 1820, 1996.

Hill, T.W., Inertial limits on corotation, *J. Geophys. Res.,84*, 6554, 1979.

Hill, T.W., et al., in *Physics of the Jovian Magnetosphere*, edited by A. J. Dessler, Cambridge University Press, 1983.

Hill, T.W., and D.H. Pontius, Plasma injection near Io, submitted to *J. Geophys. Res.*, July 1997.

Hinson, D.P., et al., Jupiter's ionosphere: Results from the first Galileo radio occultation experiment, *Geophys. Res. Let.,24*, 2107, 1997.

Huang, T.S., and T.J. Birmingham, The polarization electric field and its effects in an anisotropic rotating magnetospheric plasma, *J. Geophys. Res.,97*, 1511, 1992.

Huang, T.S., and T.W. Hill, Corotation lag of the jovian atmosphere, ionosphere and magnetosphere, *J. Geophys. Res.,94*, 3761, 1989.

Huddleston, D. E., et al., Ion cyclotron waves in the Io torus during the Galileo encounter: Warm plasma dispersion analysis, *Geophys. Res. Let.,24*, 2143, 1997.

Huddleston, D. E., et al., Ion cyclotron waves in the Io torus: Wave dispersion, free energy analysis, and SO_2^+ source rate estimates, sumbitted to *J. Geophys. Res.*, 1998.

Hui, C-H., and C.E. Seyler, Electron acceleration by Alfvén waves in the magnetosphere, *J. Geophys. Res.,97*, 3953, 1992.

Kaiser, M. L., Observations of non-thermal radiation from planets, in *Plasma waves and instabilities at comets and in magnetospheres*, AGU Geophysical Monograph 53, Tsurutani and Oya ed., 1989.

Kennel, C.F., and F. Engelmann, Velocity space diffusion from weak plasma turbulence in a magnetic field, *Phys. Fluids, 9*, 1966.

Kennel, C.F., and H.E. Petschek, Limits on the stably trapper particle fluxes, *J. Geophys. Res.,71*, 1, 1966.

Lazarus, A. J. and R. L. McNutt, Low-energy plasma ion observations in Saturn's magnetosphere, *J. Geophys. Res.,88*, 8831, 1983.

Leblanc, Y., G. A. Dulk, and F. Bagenal, On Io's excitation and the origin of Jupiter's decametric radiation, *Astron. and Astrophys., 290*, 660L, 1994.

Lecacheux, A., et al., Spectra and complete polarization state of two, Io-related, radio storms from Jupiter, *Astron. Astrophy. 251*, 339, 1991.

Lee, L. C., Theories of non-thermal radiation from planets, in *Plasma waves and instabilities at comets and in magnetospheres*, AGU Geophysical Monograph 53, Tsurutani and Oya ed., 1989.

Lee, M. A., Coupled hydromagnetic wave excitation and ion acceleration upstream of the Earth's bow shock, *J. Geophys. Res.,87*, 5063, 1982.

Lysak, R., et al., Formation of double layers on auroral field lines, in *Symposium on Plasma Double Layers*, Michelsen and Rasmussen, ed., RisøNational Laboratories, 1982.

Linker, J.A., M.G. Kivelson, and R.J. Walker, An MHD simulation of plasma flow past Io: Alfvén and slow mode perturbations, *Geophys. Res. Let.,15*, 1311, 1988.

Linker, J.A., M.G. Kivelson, and R.J. Walker, A three-dimensional simulation of the plasma past Io, *J. Geophys. Res.,96*, 21053, 1991.

Kliore, A. J., et al., The atmosphere of Io from Pioneer 10 radio occultation measurements, *Icarus, 24*, 407, 1975.

Kletzing, C.A., Electron acceleration by kinetic Alfvén waves, *J. Geophys. Res.,99*, 11095, 1994.

Kivelson, M.G., et al., Io's interaction with the plasma torus: Galileo magnetometer report, *Science, 274*, 396, 1996.

Kim, S. J., et al., Images of aurorae on Jupiter, *Nature, 353*, 536, 1991.

McGrath, M. A., and R. E. Johnson, Charge exchange cross sections for the Io plasma torus, *J. Geophys. Res.,94*, 2677, 1989.

Machida, S., et al., The electromagnetic ion cyclotron instability in the Io torus, *J. Geophys. Res.*,*93*, 7545, 1988.

Mälkki, A., et al., A statistical survey of auroral solitary waves and weak double layers 1. Occurance and net voltage, *J. Geophys. Res.*,*98*, 1521, 1993.

Menietti, J.D. and J.L. Burch, DE 1 particle and wave observations in an AKR source region, in *Auroral Plasma Dynamics*, ed. R.L. Lysak, 1993.

Menietti, J.D. and D.A. Gurnett, Whistler propagation in the Jovian magnetosphere, *Geophys. Res. Let.*,*7*, 49, 1980.

Meyer-Vernet, N., et al., Temperature inversion in the Io plasma torus, *Icarus, 116*, 202, 1995.

Moncuquet, M., Équilbre et confinement du tore de plasma d'Io dans la magnétosphère de Jupiter: observations d'Ulysse et modélisation, Ph.D. thesis, Université Paris, 1998.

Neubauer, F.M., Nonlinear standing Alfvén wave current system at Io: Theory, *J. Geophys. Res.*,*85*, 1171, 1980.

Piddington, J.H., and J.F. Drake, Electrodynamic effects of Jupiter's satellite Io, *Nature, 217*, 935, 1968.

Pontius, D.H., The implications of variable mass loading in the Io torus: The Jovian flywheel, *J. Geophys. Res.*,*100*, 19531, 1995.

Pontius, D.H., Radial mass transport and rotational dynamics, *J. Geophys. Res.*,*102*, 7137, 1997.

Prangé, R., et al., Rapid energy dissipation and variability of the Io-Jupiter electrodynamic circuit, *Nature, 379*, 323, 1996.

Press, W. H., et al., 1992, *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press, 1992.

Queinnec, J., and P. Zarka, Io-controlled decameter arcs and Io-Jupiter interaction, submitted to *J. Geophys. Res.*, 1998.

Raadu, M.A., The physics of double layers and their role in astrophysics, *Phys. Rep., 178*, 25, 1989.

Sanbonmatsu, K. Y., et al., Non-Markovian electron diffusion in the auroral ionosphere at high Langmuir-wave intensities, *Geophys. Res. Let.*,*24*, 807, 1997.

Schamel, H., and S. Bujarbarua, Analytic double layers, *Phys. Fluids*, *26*, 190, 1983.

Scharer, J. E., and A. W. Trivelpiece, Cyclotron wave instabilities in a plasma, *Phys. Fluids, 10*, 591, 1967.

Scherb, F., and W. H. Smyth, Variability of [O I] 6300-$\mathring{A}$ emission near Io, *J. Geophys. Res.*,*98*, 18729, 1993.

Schneider, N. M., Sodium in Io's extended atmosphere, Ph.D. Thesis University of Arizona, 1988.

Schreier, R., et al., A two-dimensional model of plasma transport and chemistry in the jovian magnetosphere, submitted to *J. Geophys. Res.*, 1998.

Seiff, A., et al., Structure of the atmosphere of Jupiter: Galileo probe measurements, *Science, 272*, 844, 1996.

Shapiro, V.D., and V.I. Shevchenko, The nonlinear theory of interaction between charged particle beams and a plasma in a magnetic field, *JETP, 15*, 1053, 1962.

Shemansky, D.E., and B.R. Sandel, The injection of energy into the Io plasma torus, *J. Geophys. Res.,87*, 219, 1982.

Smith, R.A., and C.K. Goertz, On the modulation of the Jovian decametric radiation by Io: 1. Acceleration of charged particles, *J. Geophys. Res.,83*, 2617, 1978.

Smith, R.A., and D.F. Strobel, Energy partitioning in the Io plasma torus, *J. Geophys. Res.,90*, 9469, 1985.

Smyth, W. H., Neutral cloud distribution in the jovian system, *Adv. Space. Res. 12*, (8)337, 1992.

Southwood, D.J., and M.G. Kivelson, An approximate description of field-aligned currents in a planetary magnetic field, *J. Geophys. Res.,96*, 67, 1991.

Strobel, D.F., and S.K. Atreya, Ionosphere, in *Physics of the Jovian Magnetosphere*, edited by A. J. Dessler, Cambridge University Press, 1983.

Strobel, D. F., et al., On the vertical thermal structure of Io's atmosphere, *Icarus, 111*, 18, 1994.

Summers, Michael E., and Darrel F. Strobel, Photochemistry and Vertical Transport in Io's Atmosphere and Ionosphere, *Icarus, 120*, 290, 1996.

Terasawa, T., Particle scattering and acceleration in a turbulent plasma around comets, in *Plasma Waves and Instabilities at Comets and in Magnetospheres*, AGU Geophysical Monograph 53, Tsurutani and Oya, editors, 1989.

Thompson, B.J., and R.L. Lysak, Electron acceleration by inertial Alfvén waves, *J. Geophys. Res.,101*, 5359, 1996.

Tokar, R.L., et al., Light ion concentrations in Jupiter's inner magnetosphere, *J. Geophys. Res.,87*, 2241, 1982.

Vasyliūnas, V.M., A survey of low-energy electrons in the evening sector of the magnetosphere with OGO 1 anbd OGO 3, *J. Geophys. Res.,73*, 2839, 1968.

Vasyliūnas, V.M., Fundamentals of current description, *Magnetospheric Currents*, ed. Potemra, *Geophys. Monogr. Ser., 28*, 63, 1984.

Volwerk, M., Strong double layers in astrophysical plasmas, Ph.D. thesis, Utrecht University, 1993.

Volwerk, M., Radiation from electrostatic double layers in laboratory plasmas, *J. Phys. D, 26*, 1192, 1993.

Wagner, J.S., et al., A simulation study of the loss cone driven cyclotron maser applied to auroral kilometric radiation, *Radio Sci., 19*, 509, 1984.

Waite, J.H., et al., The Jovian aurora: Electron or ion precipitation?, *J. Geophys. Res.,93*, 7244, 1988.

Waite, J. H., et al., ROSAT observations of the Jupiter aurora, *J. Geophys. Res.*,*99*, 14799, 1994.

Wang, K., et al., Constraints on jovian plasma properties from a dispersion analysis of unducted whistlers in the warm Io torus, in press, *J. Geophys. Res.*, 1998.

Warnecke, J., et al., Ion cyclotron waves observed at Galileo's Io encounter: Implications for neutral cloud distribution and plasma composition, *Geophys. Res. Let.*,*24*, 2139, 1997.

Williams, D.J., et al., Electron beams and ion composition measured at Io and in its torus, *Science, 274*, 401, 1996.

Wilson, J. K., Io's fast sodium clouds and the escape of Io's ionosphere, Ph.D. thesis, University of Colorado, 1996.

Wolf-Gladrow, D.A., F.M. Neubauer and M. Lussem, Io's interaction with the plasma torus: A self-consistent model, *J. Geophys. Res.*,*92*, 9949, 1987.

Wright, A.N., The interaction of Io's Alfvén waves with the Jovian magnetosphere, *J. Geophys. Res.*,*92*, 9963, 1987.

Wright, A.N., and S.J. Schwartz, The transmission of Alfvén waves through the Io plasma torus, *J. Geophys. Res.*,*94*, 3749, 1989.

Wright, A.N., and S.J. Schwartz, The equilibrium of a conducting dody embedded in a flowing plasma, *J. Geophys. Res.*,*95*, 4027, 1990.

Wu, C.S., et al., Theory of decametric radio emissions from Jupiter, *Icarus, 18*, 192, 1973.

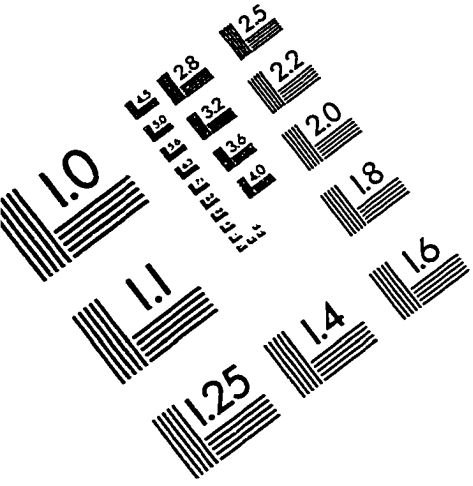
Wu, C.S., and L.C. Lee, A theory of the terrestrial kilometric radiation, *Astrophys.J., 230*, 621, 1979.

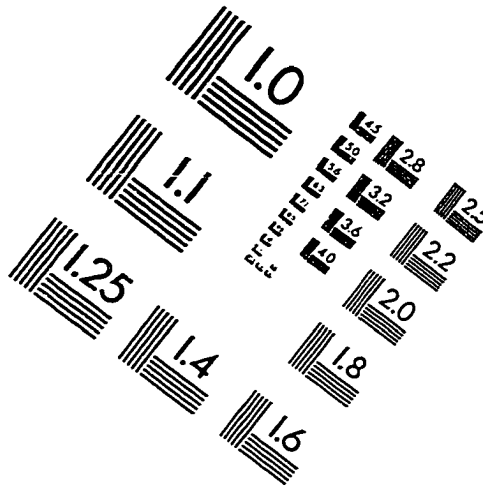
Zarka, P., Beaming of planetary radio emissions, in *Planetary Radio Emissions II*, ed. Rucker et al., 1988.

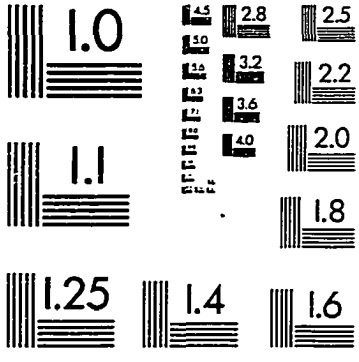
Zarka, P., The auroral radio emissions from planetary magnetospheres: What do we know, what don't we know, what do we learn from them?, *Adv. Space Res., 12*, (8)99, 1992.

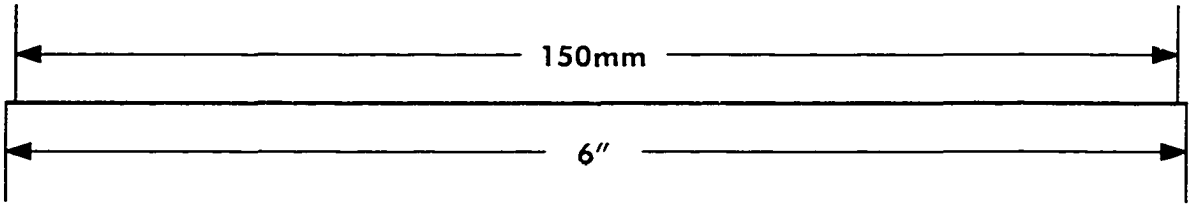
Zarka, Philippe, et al., A scenario for Jovian S-bursts, *Geophys. Res. Let.*,*23*, 125, 1996.

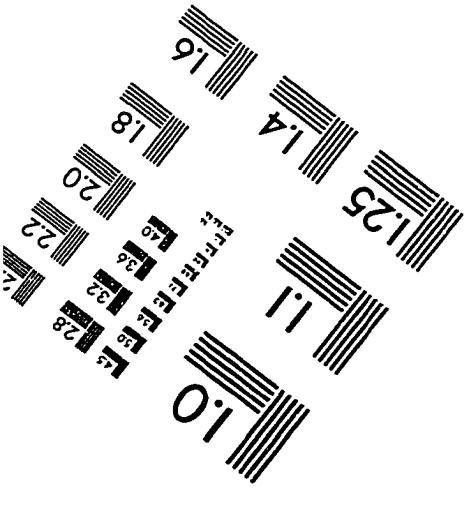
IMAGE EVALUATION TEST TARGET (QA-3)

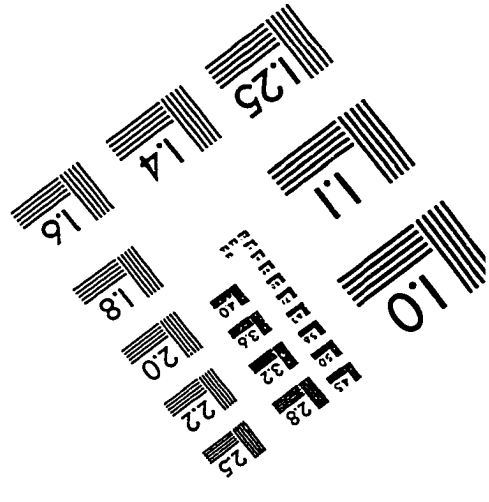












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

IN THE WORK OF THEIR HANDS IS THEIR PRAYER: LABOR AND CULTURE ON THE AMERICAN FRONTIERS, 1830-1930

by

JOEL FRANCIS DAEHNKE

B.A., Montana State University, 1986

M.A., Northern Arizona University, 1990

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of English

1998

This thesis entitled:
In the Work of Their Hands is Their Prayer: Labor and Culture on the American Frontiers, 1830-1930
written by Joel Francis Daehnke
has been approved for the Department of English

Mary K Klages

Dr. Mary Klages

Lee Krauth

Dr. Lee Krauth

Date 4/28/98

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Daehnke, Joel Francis (Ph.D., English)

In The Work of Their Hands is Their Prayer: Labor and Culture on the American Frontiers, 1830-1930

Thesis directed by Associate Professor Mary Klages

In the Work of Their Hands is Their Prayer: Labor and Culture on the American Frontiers, 1830-1930, examines a wide variety of written responses to westward expansion, paying particular attention to the manner in which these texts embody cultural attitudes and values with respect to the relationship between human activity, and the transformation of the American Landscape. The dissertation argues that within the Protestant, and Republican ideologies which lent redemptive status to the work of expansion on the American frontiers, lay also a set of assumptions about the nature of human effort; the transformation of the western landscape, in other words, raised cogent issues with regard to the morality, efficacy, and ends of human activity as it shaped, and was in turn shaped by, the natural environment. In examining the utopic character of both the Protestant vision of America as a New Canaan, and the Classical Republican ideal of the city state, (each of which, and often in a complex relationship with one another, influenced American thought during this period) I suggest that beyond the redemptive status each claimed for the frontiers, there existed a fundamental discourse which sought the redemption of human labor itself. Reclaiming the frontiers for civilization, I argue, essentially held out the hope for an idealized transformation of the human condition.

The dissertation examines the redemptive discourse and imagery inherent in the representation of westward expansion by examining models

of frontier production and consumption in Caroline Kirkland's *A New Home, Who'll Follow?*; representations of the profane and sacred landscape in early writings on Yellowstone National Park; the "redeemed" culture of frontier leisure, in an examination of the textual representation of the sport of fly fishing; and finally, a chapter on Willa Cather's *Death Comes for the Archbishop* examines the imagery of the physical body in the context of Christian grace and redemption.

For my Father and Mother

ACKNOWLEDGMENTS

This project has only come about through the help, guidance, and patience of a number of people whom I would like to thank. My committee, each of whom contributed thoughts and perspectives which shape my understanding and writing, have all been more than gracious. Mary Klages, warm, and encouraging; Lee Krauth, patient with me in all things; Mike Preston, my first supporter, and maker of space in which to work; Adeleke Adeeko, latecomer, but insightful as anyone, and Patricia Nelson-Limerick, whose willingness to read my work was flattering, to say the least. To all of these people, I owe my best thanks.

I would also like to thank David Simpson, who helped so much in giving me a way to read, and who helped develop in me a sense of scholarship, and a hopeful sense of myself as a scholar. The Graduate Directors who, during my time here have found ways to support me: Eric White, John Stevenson, Beth Robertson, each of whom, in their own way, reflect the support of this English Department at the University of Colorado, and who have made me feel both welcome, and appreciated here.

I would like also to acknowledge the gracious support of the Charlotte W. Newcombe Foundation, the Florence Husted Lowe Foundation, the Graduate School at the University of Colorado, and the Dean's Small Grant program, for the generous funding which helped to bring this work to fruition.

Lee Whittlesey, Yellowstone National Park Archivist, and the staff at the Yellowstone National Park Library in Mammoth Hot Springs, were

extremely supportive while I was doing research there; Lee was a strong reader of my work on the park.

There are those to whom an acknowledgment is a poor excuse for an apology. To Carla, Sarah, Laura, and Mollie --- patient, supportive, loving, and unconditional in all respects for eight years--- you have my love.

Finally, to the great trout of the rivers of the Rocky Mountain West, who have taught me the importance of a good presentation, a well-tied fly, and a drag-free drift, I owe a sense of perspective and balance which I simply couldn't have acquired by harassing anyone else.

CONTENTS

CHAPTER

1. Introduction: The American Frontier as the Space of Redemption in American Letters, 1830-1930.....1

2. "They Don't think it Right to Spend Money So": Cultivating a Market Ethos in Caroline Kirkland's' *A New Home, Who'll Follow?*.....18

3. "A Remarkable Load of Original Sin": Redeeming the Profane Landscape in Yellowstone National Park.....67

4. "If God Made Fishin', a Feller Orter Fish: Grace, Leisure, and the Culture of Fly Fishing.....141

5. "At Our Meanest Tasks": Redemption and the Domestic Economy in Willa Cather's *Death Comes for the Archbishop*.....162

6. Conclusion: Having Done With Calendared Time: Repose, Redemption, and the Ideology of Progress.....208

Bibliography.....211

CHAPTER 1

INTRODUCTION: THE AMERICAN FRONTIER AS THE SPACE OF REDEMPTION IN AMERICAN LETTERS, 1830-1930

> Lands, boundless in extent, exhaustless in fertility, lying under every variety of climate from the tropical to the arctic; accessible in all their parts by continuous water-courses of magnitude unparalleled on the globe, containing so much to stimulate the nobler faculties and gratify the senses; so much that is calculated to induce a high state of physical development and fine perceptions of the beautiful, the grand, and the true; lands whose primeval glory, when it shall have become ancient, will form the theme of the poet and glow on the page of the historian; though too feebly sung and written to convey to future ages what the present feels. It must be the theatre of a life larger than human prophecy can foretell![1]

These closing words to *Life in Prairie Land*, by Eliza Farnham, offer a compressed vision of what, by the 1830's, much of the American public had come to believe about the vast spaces of the western frontiers; fortunate topography, good climate, fertile soil and virile health, fine perceptions, noble faculties, and not least, the peculiar gaze of poet and historian--- all these things were drawn together in a "theatre of life" being played out on the American plains, a theatre in which any observer understood the present moment in relation to both past and future of a grand human narrative. To be sure, one sees, in this narrative, those aspects of the civilizing impulse which are the ready inscriptions of Anglo-European culture as it pursued the policies of westward expansion: the agrarian landscape, an idealized physical development, reason made "fine" in its relation to an understanding of the beautiful, and of course, the written text as testament to this whole project--- past, present, and future. Though Farnham feels, perhaps, that both prophecy and history are likely to fail the fullness, the raw experiential nature of the present

moment, it remains clear that whatever unfolds on the prairies of the American Republic links past to future via a present that is, for all intents and purposes, a *presence*, an embodied historical consciousness making time out of broad vistas, and conversely, making broad vistas out of the sweep of time. The expansive present, the fullness of feeling that goes into such passages, comes to be felt as a sense of the temporal which has become overdetermined, as it were, by the sheer potential of landscape understood *as* theatre, as plot, as written narrative. Wherever this *presence* fixes its gaze, there a specific story unfolds. Whomever this *presence* is, defines narration, perspective, the Subject, and the object, of history.

This dissertation takes up the *presence* of Anglo-European culture as it laid claim to the American Frontier in an act of narration borne of a complex, and often ruthless understanding of historical process. We would seem to be at a point in which the state of Western American studies ought to allow us to make ready distinctions between this concept of historical narrative as a specific discourse whose ideological determination is in clear relation to structures of power, and "history" understood as the continuity and discontinuity of actual processes, events, modes of production, occupation of space, etc.; in other words, all those things which go into the material conditions of life, and which we should all agree constitutes some sort of objective reality that *is* the past, present, and ultimately, the future.[2] This may be an optimistic statement, and yet I mean to make it clear that in this dissertation, "narration," and "narrative" imply what Fredric Jameson has described as a network of "socially symbolic acts," which, understood as a particular way of viewing historical processes, are laden with the ideological structures of specific

relations of production, types of exchange, geo-political world-views, etc., all of which are historically conditioned, yet fundamentally reducible to what he understands to be the essential concern of history itself: "the unity of a single great collective story... the collective struggle to wrest a realm of Freedom from a realm of Necessity."[3] Narration, and narrative, then, are inherently inscribed with a teleological concept of historical process, which places particular value on the ideology of progress, of continuity and design in those processes, and ultimately, a particular and privileged "Subject" of history whose "Freedom" comes to be seen as the true "end" of human development.

However, it should be noted that there is a clear assumption in Jameson's articulation of a fundamental and "unified" impulse driving historical processes which inscribes the concepts of struggle, of Necessity, and Freedom within a particular oppositional relation between human beings and the natural world in which they live. Necessity, in other words, tends to be bound up with an aspect of human experience which finds itself engaged in a struggle to wrest the most basic human needs from an inhumane natural world; "collectivity," as a result, comes to define a sort of shared opposition to nature which inherently transcends cultural, political, gendered, and ethnic boundaries--- boundaries which come to be seen as only marginally divisive so long as all human effort shares this fundamental concern, and so long as this concern fashions the terms of historical understanding. The problem lies not so much within a materialist theoretical perspective per se; it should be clear to anyone that the life process, and therefore the condition of history itself, is undeniably rooted in the fact that human beings inhabit the world, and draw on its resources for their own maintenance--- this *is* most definitely a

shared aspect of the human condition. Rather, I would simply argue that Jameson's assertion of the collective nature of the struggle for Freedom tends to neglect the consequences for history of the privatization of material resources drawn from the natural world, which fact too, is undeniable, yet which has clearly come to inscribe historical narrative with an element of divisiveness that begins to mediate any relations between human beings and the natural environment. Marx himself, with no particular reverence for the natural world, was nonetheless aware of this "social" dilemma at the heart of the human relation to the natural world, and he reflected that: "[a]t the same pace that mankind masters nature, man seems to become enslaved to other men or to his own infamy."[4] What one comes to understand, then, is that wherever the natural world comes to constitute the primary obstacle in a narrative of historical process, the realities of social conflict, political struggles, ethnic strife, etc., all tend toward an alarming erasure by virtue of their secondary status, or more to the point, are subsumed under a larger "collective" category in which historical "progress" becomes idealized as a mastery over nature rather than a mastery of one group of individuals over another.

And yet it would be unfair to argue that, while the articulation of a "collective," and "unified" struggle constitutes a grave misunderstanding of the reality of historical process, it therefore has no real history of its own; the redemptive, and teleological elements inherent in this historical world-view have, after all, found their way into manifold narrative structures in the realm of what Jameson has referred to as "socially symbolic acts"--- culturally and historically specific instances of cultural

production whose interpretation, he suggests, turns on the restructuring of:

> the problematics of ideology, of the unconscious and desire, of representation, of history, and of cultural production, around the all-informing process of *narrative*,... the central function or *instance* of the human mind.[5]

While it needs to be understood that the construct of a redemptive, and collective struggle to wrest Freedom from a naturalized realm of Necessity does indeed tend toward an elision of the history of conflicts which constitute the real, and social context of a more rational historical world-view, it remains the fact that historically, the articulation of redemptive narratives has been laden with rhetorical strategies whose ideological value is *precisely* such an elision; in other words, wherever, and whenever redemptive narratives make themselves felt as key elements of the cultural production of texts, discourse, images, etc. within a specific cultural context, they tend, at the expense of alternative world-views, toward a valorization of a unique social vision, the value of whose attainment is constructed as a universalized good. Historically, the civilizing impulse of Anglo-European culture, its religious, political, aesthetic, and ethical world-view has operated in the North American landscape as a set of assumptions about the innate supremacy of its own practices and behaviors; the erasure of alternative cultural perspectives has been largely pursued as a program which, quite aggressively, asserts the inherent value of assimilation, mastery over natural resources, Judeo-Christian redemption, capitalist modes of production, etc., all as a universally accepted, and morally superior world-view. Moreover, as the landscape comes to be remade in the image of the universal good, historical processes come more and more to appear as redemptive in

purpose, and essentially driven by the "rational" pursuit of a "shared" goal.

This dissertation takes as its focus that most prescient narration of historical process on the American frontiers, Manifest Destiny, and its manifold articulations which have come to constitute a written *oeuvre* ranging broadly through literary texts, journals of exploration, sermons, laws and written policy, promotional literature, diaries, travel literature, and so on. Anders Stephanson has suggested that the doctrine of Manifest Destiny, at its core, constitutes a "peculiar fusion of providential and republican ideology that took place after the Revolution":

> Visions of the United States as a sacred space providentially selected for divine purposes found a counterpart in the secular idea of the new nation of liberty as a privileged "stage"... for the exhibition of a new world order, a great "experiment" for the benefit of humankind as a whole.[6]

Both Christian and republican ideology inscribe themselves ambitiously, I would argue, into a progressive scheme which charts the distance between a realm of Necessity and a realm of Freedom as a sort of collective journey toward a specific, and valorized vision of a social community, a community which, in turn implies a determined, and valorized set of conditions for individual life within that community. This dissertation examines the manner in which these redemptive narratives found their way into a variety of writings on the American frontiers, reflecting on both social relations and individual experience as it unfolded in a unique, and diverse physical environment. At the heart of both the Christian and republican redemptive schemes, I would like to suggest, lies a construct of an "exemplary" physical space which sought to create the conditions of Freedom as a set of culturally determined ideas shaped by attitudes

towards nature, acceptable modes of production, types of human activity, (moral, or immoral, productive, or "unproductive" savage, or civilized) and not least, the idealized appearance of the redeemed landscape as a sort of eschatological denouement at the "ends" of history. The North American Continent itself had been early constructed as the exemplary site for the eschatological designs of Providence; in many ways the articulation of Manifest Destiny is a continuation of that pattern of representation.[7] Such a ground is at once spatial and temporal, insofar as the idea of an "elect" landscape suited to the designs of specific historical processes is fraught with determinist notions of its past, present and future, as well as the types of human activity which unfold there. The ideology of Manifest Destiny is itself unthinkable apart from an historical world-view which sees little difference between the conquest of space, and the chronology of an expansive enterprise.[8]

It seems clear, then, that while the policy of expansion guiding westward migration was ideologically a redemptive project, one which sought to alter the face of the landscape in the making of a "proper" ground for Anglo-European culture, representations of the American frontiers had also to contend with constructs of human activity as a primary agency in the remaking of landscape. In other words, redemptive narratives are inherently bound up with a transformative discourse on the notion of physical activity itself; Christian redemption, charting the body's path from fallen nature to a state of grace, is writ large with an ideology of human strife transformed to repose, and the ultimate denial of the physical. Classical Republicanism, too, charts a redemption of physical existence in the ideal of the virtuous citizen, whose freedom from mundane labor creates the civic space in which the performance of public

deeds and actions assures the continuity of individual life through a collective remembrance that outlives the physical body.[9] The "peculiar fusion" of providential and republican ideologies at the core of Manifest Destiny makes it difficult to separate the notion, of, say Christian virtue, from a concept of civic virtue, or a concept of the life of toil from the cursed nature of a fallen human condition. And yet, what remains important here, is the notion that the ideology of redemption inherent in this complex relationship ultimately suggests that human activity must, irreducibly, create the conditions of its own Freedom. Providence may have withheld the North American continent for the unfolding of Christian eschatology, or the wealth of nature held forth its bounty for the creation of the great Republic; yet however complex the understanding of historical processes at work in an expansive enterprise, a broad array of human activities remained at the core of its realization, as well as at the "ends" of its vision, however idyllic or utopian.

This dissertation, then, attempts to understand the redemptive impulse at the heart of representations of the American frontiers, not simply through an examination of a transformative discourse bound up with the alteration of landscape, but also through its implied transformation of specific practices and behaviors, whose place in that landscape is defined by a range of activities which are themselves heavily invested with cultural meaning. While I understand that the notion of "activity" both connotes and denotes a wide range of human experience, I have sought to categorize my use of this terminology primarily in accordance with what Hannah Arendt, in *The Human Condition*, has argued is a comprehensive, and hierarchical typology of human activity, which I will try to summarize here briefly. Arendt's argument, itself

largely dependent on Classical Republican theory, divides human activity into four arenas of behavior, each of which can be constructed with a view to the physical body by means of a typological, universal figure engaged in specific behaviors. Hierarchically situated as behaviors or activities which themselves form an implicit ideology of a redemptive path leading from necessity to freedom, she suggests that the condition of "labor," which is utterly, and irreducibly bound up with the continuity of the life processes, is the fundamental proposition of life, yet insofar as it is transient, and privatized within the needs of the body, hardly to be conceptualized as "human." The type of individual caught at this level of existence, indeed, is considered at a level of nature, as *animal laborans*. Arendt distinguishes "labor" from what she considers to be "work," an arena of activity born of the capacity of labor to produce a surplus of human energy, made possible at those levels of civilization which have managed to transcend subsistence as a full-time consumption of time and energy. The activity of "work," for Arendt, constitutes the experiential realm of *homo faber*, the individual whose productive capacity is bound up the construction of durable spaces and artifacts which ultimately "shelter" life from the natural, world, and create a world fit for human habitation. "Action," Arendt asserts, constitutes the true realm of lived experience as a set of social relationships, and is most easily typologized in the concept of the "citizen" whose behavior in a social world constructed by *homo faber* is born of the need for self-realization, the need to distinguish one's self through public excellence and deeds which outlast the human body itself. Already, a certain level of freedom is implied in the conditions of the citizen, whose freedom from the daily concerns of life allow him to get beyond the privatized life of the household and its attention to physical

need. Finally, Arendt suggests an attitude of withdrawal in the concept of the *vita contemplativa*, which is essentially a valorized state of repose in which the individual, utterly detached from worldly concerns, is free to meditate on a realm of universals in a virtual transcendence of desire, and the endless cyclicality inherent in the natural world. Again, these categorizations of human activities are radically hierarchical in their relation to one another, and constitute a redemptive schema at the heart of *The Human Condition.*[10]

This is an overgeneralization of Arendt's construct of human activity to be sure. Nor is it my intention here to take up a critique of her argument, which has strong implications not only as an implicit valorization of certain patterns of cultural development, but which also, as she is well aware, suggests conditions for the appropriation of physical space, labor, and ultimately, the conquest of others. Rather, what I would mean to argue is the utility of this categorization in the examination of representations of types of human activity as they appear in texts dealing with the American frontiers; in other words, if Manifest Destiny, as a complex understanding of historical process which made its way into textual representations of the frontiers, is itself inscribed with a progressive ideology of human activity in the remaking of the American landscape, then Arendt's typology ought to offer a means to deconstruct the cultural forces driving this expansive project. Indeed, William Gilpin's articulation that the "*untransacted* destiny of the American people is to subdue the continent" is a virtual catalog of the physical body at work in the reconstruction of the landscape, as the American finds virtue in the will:

> to rush over this vast field to the Pacific Ocean--- to animate the many millions of its people, and to cheer them upward... --- to agitate these herculean masses--- to establish a new order in human affairs... --- to regenerate superannuated nations--- ... to stir up the sleep of a hundred centuries--- to teach old nations a new civilization... to confirm the destiny of the human race--- to carry the career of mankind to its culminating point--- to cause a stagnant people to be reborn--- to perfect science--- to emblazon history with the conquest of peace--- to shed a new and resplendent glory upon mankind--- to unite the world in one social family-- to dissolve the spell of tyranny and exalt charity-- to absolve the curse that weighs down humanity, and to shed blessings around the world![11]

Rushing, agitating, regenerating, awakening, carrying, shedding, all this terminology of a moral agency at work in expansion is writ large in the images of physical exertion, and its implications for the "collective struggle to wrest a realm of Freedom from a realm of Necessity" are clear enough.

This dissertation, then, seeks to examine some aspects of the relationship of the individual body to the national body, so to speak. I have chosen, perhaps because of the pervasiveness of Christian imagery in the construct of the frontier as a contested ground between the redeemed, and the unredeemed landscape, to consider in terms of The Christian pattern of salvation the ways in which textual representations meditate on the notion that the redemption of the frontier through human activity creates the conditions for the redemption of the human condition itself. The Christian concepts of fallen nature, the curse of labor, works, grace, and finally, the millennial vision of a redeemed and humanized landscape as a sort of earthly paradise, all have their place in the ideology of Manifest Destiny.

And yet, I have also sought to acknowledge the prevalence of what Mark Hanley has suggested was a growing concern on behalf of Protestantism that a secular republican ideology might create a "rising message of popular material redemption that had nothing to do with religion."[12] For this reason, I also choose to examine frontier texts with a respect to a concern that the terminology and values inherent in the Christian redemptive scheme might be co-opted by a strictly secular, and temporal view of progress.

Finally, I would like to suggest that the teleological impulse inherent in narratives of redemption implies not only a progressive view of historical processes, but also, inscribes in the space between Necessity and Freedom, uncivilized and civilized, profane and sacred, labor and the end of labor, etc., the forms of desire that come to define the nature of historical struggle itself, so to speak. Wherever redemptive narratives find articulation in the broadest realms of cultural production, they are charged not only with the representation of the images of redemption, but with the images of desire as a motive force driving individual and social goals. As a result, the sweep of redemption understood as a culmination of historical processes can be viewed as a progressive erasure of desire, whether it make itself felt as a physical, social, or indeed, a spiritual concept of "need," and may be felt at the level of the individual, or of society. Again, this erasure is also an erasure of social conflicts which go into the pursuit of specific cultural goals, and its ideological component is essentially at the heart of this dissertation's examination of redemptive narratives. For the purpose of examining the ideological aspects of this

erasure on the American Frontiers, I'd like to suggest that I find its representation most cogently represented in four distinct areas of imagery inherent in frontier literature, and my choice of textual material for this project has been defined as follows:

1. The arena of production.

This first chapter, on Caroline Kirkland's *A New Home, Who'll Follow* discusses this text's use of Edenic imagery in a fantasy of the frontier garden as a showcase of diverse consumption and women's productivity. Ultimately, I argue that Kirkland's critique of frontier patterns of work and consumption reveal her distaste for the lack of a vital market economy in 1830's Michigan. "Desire," in this area of representation, comes to be seen as the problematic of commodity scarcity, the lack of wage laborers to provide services, and the lack of a stable domestic space made comfortable by both goods and services. Kirkland's project of "cultivating" on the Michigan Frontier a solid work ethic, and an interest in the purchase of commodities is, in this instance, both the evocation of commodity desire in her neighbors, and for Kirkland, the personal fantasy of the erasure of that desire as goods and labor become available to the woman of means.

2. The frontier landscape.

A chapter follows on the early writings of Yellowstone National Park; tracing the works of explorers, authors, and visitors, the chapter seeks to understand the development of Yellowstone as a nationalist project in the reclamation of a "profane" landscape for a growing, and "moral" culture of tourism. The idea of Yellowstone as a landscape both profane and sacred

is essentially bound up with an aesthetic view of its remarkable landscape. While many features of the park have tended to evoke the aesthetic discourse of a traditional sublime, or picturesque definition of pleasing landscape, Yellowstone's thermal features have occasioned a consistent discourse which has been bound up with both images of the profane, as well as the imagery of incessant labor at work in geological forces. Thermal activity, insofar as it appears to be bound up primarily with the endless generation and expenditure of natural force, essentially comes to be viewed as a landscape of labor. This chapter examines desire as the project for the reclamation of the profane, and laboring landscape for a burgeoning culture of leisure. The erasure of desire, I would suggest, finds its culmination in the detached, objective gaze of the tourist who, in the midst of holiday, is free to play in the park, and who, in addition, has made of the laboring landscape a site for his or her own relaxation.

3. The physical body.

A chapter on the literature of fly fishing in the west discusses frontier leisure in the context of Christian Grace (I argue here that leisure, or free time, constitutes the redemptive element in a secular narrative of progress). This chapter continues a discussion of leisure culture within the context of the physical body as the site in which to examine the erasure of desire. I argue that the codified practices and ethical constraints on productivity inherent in the sport of fly fishing draw on the wasteful, and unproductive expenditure of time as a means to distance the physical body from a concept of desire bound up with the notion of appetite. Making of the history of fishing a process of radical withdrawal

from its origins in necessity and subsistence, I suggest that fly fishing aestheticizes the body by virtue of a discourse on physical grace, and further, redeems it by an ethics of self-restraint and "under-production"; the erasure of desire, in this context, is bound up with the denial of physical need, and the rejection of appetite.

4. The spirit.

A chapter on Willa Cather's *Death Comes for the Archbishop* examines the erasure of desire within the realm of spiritual and cultural production, as defined by high European culture. Desire is bound up clearly, with the hopes of religious salvation for both the individual and the inhabitants of the American Southwest of the mid-nineteenth century, and yet too, there is the desire for the production of religious/artistic artifacts which will provide a legacy and "continuity" for a life lived on earth. The erasure of desire finds its culmination within the valorization of the religious/aesthetic life of contemplation, in which the bodily aspect of repose, and the pure, idealized consumption of art and high culture suggest a blend of Christian grace and high cultural refinements.

What follows here is not in any way meant as an historical meditation on the development of an idea. Redemptive schemes, from the point of view of this dissertation, are constantly recycled, put back into play, and re-situated with respect to a diverse topography of landscape, and shifting grounds of human activity. The historical period under consideration here hardly saw a fruition of either Christian or republican processes. Rather, my intentions are to examine frontier texts as they meditate on specific aspects of redemptive narratives. There is, I suppose, an ideology of progress inherent in the manner in which these chapters

reflect on the fantasy of the Edenic as a ground of diversified consumption and productivity in Caroline Kirkland's *A New Home, Who'll Follow?*, the redemption of the fallen, and "laborious" landscape of Yellowstone for a growing culture of leisure, the idea of fly fishing as a construct of leisure fraught with the ideology of Christian Grace, and finally, the image of the Southwestern domestic as the space of the contemplative life in Willa Cather's *Death Comes for the Archbishop*. And yet, each of these chapters, I would argue, contains the imagery of an expansive historical consciousness which sees, in the vistas of the American frontiers, and in the types of activity which unfold there, both past and present, both inception and culmination of historical process as redemptive. Ultimately, these texts contain the inscription of desire, a sort of collective "prayer" that human effort be rife with meaning, and the human body, a force in the landscape actively creating the conditions of its own Freedom.

---

[1] Eliza W. Farnham, *Life in Prairie Land*. (Urbana, IL: University of Illinois Press, 1988), p. 268.

[2] The project of rethinking the written history of the west as a problematic narrative is well underway; See, e.g. Richard White, *It's Your Misfortune and None of My Own: A New History of the American West*. (Norman, OK: The University of Oklahoma Press, 1991), and Patricia Nelson Limerick, *The Legacy of Conquest: The Unbroken Past of the American West*. (New York: W.W. Norton & Co., 1987). See also Donald Worster, *Under Western Skies: Nature and History in the American West*. (New York: Oxford University Press, 1992). For an examination of the ideological character of American literary texts with respect to the frontier, see, esp. Annette Kolodny, *The Lay of the Land: Metaphor as Experience and History in American Life and Letters*. (Chapel Hill: University of North Carolina Press, 1975), and *The Land Before Her: Fantasy and Experience of the American Frontiers, 1630-1860*. (Chapel Hill: University of North Carolina Press, 1984). See also Richard Slotkin, *Regeneration Through Violence: The Mythology of the American Frontier, 1600-1860*. (Middletown, CT: Wesleyan University Press, 1973).

[3] Fredric Jameson, *The Political Unconscious: Narrative as a Socially Symbolic Act*. (Ithaca, NY: Cornell University Press, 1981), p. 19.

[4] Karl Marx, "Speech at the Anniversary of the People's Paper," April 1856. This quotation is an opening epithet to Alfred Schmidt, *The Concept of Nature in Marx*. (London: NLB, 1971). Schmidt's text, while sensitive to the criticism of Marx 's views on

the "mastery" of the natural world, nonetheless makes a strong case for the "social" realities and conflicts at the heart of the human relation to nature, and as such, offers a good corrective to Jameson's "collective" narrative.

[5] Jameson, *Political Unconscious*, p. 13.

[6] Anders Stephanson, *Manifest Destiny: American Expansion and the Empire of Right.* (New York: Hill and Wang, 1995), p. 5.

[7] See especially, Perry Miller, *Errand Into the Wilderness.* (Cambridge, MA: Harvard University Press, 1956), and Sacvan Bercovitch, *The American Jeremiad.* (Madison, WI: The University of Wisconsin Press, 1978). Bercovitch's work, with it's attention to the pervasiveness of the concept of "mission" in the American national consciousness, is particularly of note here.

[8] Here, the idea of *translatio imperii* is a provocative way to think the historical consciousness in terms of a spatial/temporal construct. He suggests that the concept was an "agreeable [for the American consciousness of the time] double notion that civilization was always carried forward by a single dominant power or people and that historical succession was a matter of westward movement." See Stephanson, *Manifest Destiny*, p. 18. See also Henry Nash Smith, *Virgin Land: The American West as Symbol and Myth.* (Cambridge, MA: Harvard University Press, 1950), pp. 15-48 for a good discussion of the "westward course of empire" as a motive ideology driving American policy with regard to expansion.

[9] My understanding of Classical Republican theory is heavily indebted to J.G.A. Pocock's *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition.* (Princeton, NJ: Princeton University Press, 1975).

[10] See Hannah Arendt, *The Human Condition.* (Chicago: The University of Chicago Press, 1958).

[11] Quoted in Nash Smith, *Virgin Land*, p. 37.

[12] Mark Y. Hanley, *Beyond a Christian Commonwealth: The Protestant Quarrel with the American Republic, 1830-1860.* (Chapel Hill: University of North Carolina Press, 1994), p. 84.

CHAPTER 2

"THEY DON'T THINK IT RIGHT TO SPEND MONEY SO": CULTIVATING A MARKET ETHOS ON THE MICHIGAN FRONTIER IN CAROLINE KIRKLAND'S *A NEW HOME, WHO'LL FOLLOW?*

> "I take this opportunity to declare that all the naughty and unpleasant people-- all the tattlers and mischief makers,-- all the litigious,-- all the quarrelsome, -- all the expectorant, -- all the unneat, -- all the unhandsome,-- have emigrated to Iowa, Wiskonsan, (*sic*) or Texas, or some other far distant land, to this deponent unknown; and that there is not-- *meo periculo*, one single specimen of any of these classes remaining in this wide peninsula."[1]

So wrote Caroline Kirkland in the opening pages of *Forest Life*, her second collection of essays and sketches detailing life on the Michigan frontiers and her residence in the small village of Pinckney from 1835-1843. Exorcising this cast of unflattering characters as a pretext to a second work on western lifestyles, she had obviously felt the need to respond to what to her must have been the most immediate critical reception of her earlier *A New Home, Who'll Follow*, that of her Pinckney neighbors. For that text had dealt rather freely with habits and customs of hinterland society which Kirkland found onerous to her urban sensibilities and middle-class ideals of comfort--- patterns of behavior which had not only disturbed her notions of human character, but which had placed her on the margins of an economic community whose networks controlled most of the goods and services necessary for life outside of a well-developed market society. As a result, a text which had read to its *Knickerbocker* reviewer as "the most natural, pleasant and entertaining... we have read in a twelfthmonth," having " a keen eye for the ridiculous... a ready perception of the burlesque, and with a remarkable knowledge of

character" she found judged more harshly within the tight circles of rural village life.[2] Though she had hidden her authorship under the narrative guise of a Mrs. Mary Clavers, had created a fictional village named Montacute, and had apparently never suspected that copies of the book might find their way back to Michigan, Kirkland soon discovered, as Mary Kelley has suggested of other nineteenth-century women authors, that "anonymity related more to a state of mind than a state of reality."[3] As a result, she found herself facing squarely the ire of neighbors who had managed to pierce the fictional constructs of place and character to see in *A New Home* "the deep damnation of their taking-off."[4] And while in *Forest Life* Kirkland had continued to justify her representations of Pinckney inhabitants by suggesting that characterization in the book was merely the necessary foreground to a more fundamental depiction of "landscape," which, "however true its outline... is only a study for the artist unless something human appear... to give an air of life to the scene," (*FL,* 4) her neighbors appear to have understood well enough that landscape, as such, comes on the scene of representation heavily charged with the topography of social relations. To be sure, Kirkland's narrator had been reluctant to devote overmuch space in the text to a scenery as tiresome for "the sameness of [its] oak openings"[5] and mud-holes as it was appealing in its diversity of wildflowers. Rather, what any reader of the text must have discovered was that landscape, in *A New Home*, constituted most often the social geography of close living among village inhabitants in a community where the "diminutive size" of log cabins and "habits of society which allow[ed] the maid and her mistress to do the honors [of social visits] in complete equality" (*ANH,* p.4) had severely displaced Kirkland's more urban sensibilities. For if authors such as

Irving could become apprehensive of the engulfing expanse of western vistas, the frontier was for Kirkland primarily a place where cramped domestic spaces threatened the breakdown of privacy and social order.[6] At any rate, Pinckneyites were hardly amused with Kirkland's social satire and comic characterizations, and as late as 1843 a New York lawyer visiting the area had found that "she was in bad odure with her neighbors."[7]

It is important, then, to view Kirkland's *Forest Life* response as an attempt to repair the social damage of her earlier text, and precisely so given the difficulties attendant upon the loss of goodwill in a society where , as she noted in *A New Home,* the viscidities of life "may throw you entirely upon the kindness of your humblest neighbor" (*ANH*, p. 65). However, this apology and subsequent removal to literary hinterlands of all "classes" unflattering to Pinckney circles was in itself only characteristic of the ways in which *A New Home* had structured its own social satire, a satire rooted deeply in class distinctions between backwoods observed and urbane observer, between the figures of literary representation on the one hand, and an implicit alliance between narrative voice and a readership comprised of "friends back in the settlements" (*ANH,* p. 3) on the other. For although she had, "*meo periculo,* " agreed to acknowledge as an audience those with whom she must hold daily social concourse, the biting tone of the passage manages to retain this audience as objects of representation for a broader reading public in the east. In fact, so effectively does the passage speak both to and *of* its frontier readership that one comes to realize Kirkland's angry neighbors have acquired comic value *as* readers. Moreover, since

Kirkland had felt the need to school her neighbors on the aesthetics of reading landscape and character, this readership is now given its status within an ideology of uneven development where relations between author and audience range from those to whom she could freely, if at a distance, quote from the Latin, the French, and Shakespeare, to those in an immediate community whose inhabitants felt "that there are books enough in the world, and *one too many*" (*FL* 8).[8]

That Kirkland could fashion this construct of a readership whose levels of refinement and cultivation actually regressed as one moved from east to west is hardly surprising; Henry Nash Smith has argued the pervasiveness of the nineteenth-century idea that "in America one could examine side by side the social stages that were believed to have followed one another in time in the long history of the Old World.[9] However, if such a concept is taken too strongly to embody the anxiety over America's westering experience, one begins perhaps to understand Kirkland's sense of distance between herself and, to her mind, a more cultivated eastern audience, yet diminishes its potential as a *progressive* vision for the eastering gaze of the hopeful settler. To be sure, Jefferson had utilized this spatial-temporal construct as a model of socio-economic achievement, suggesting that if one only chose to begin the journey at the foot of the Rockies heading east, "in his progress he would meet the gradual shades of improving man until he would reach his, as yet, most improved state in our seaport towns."[10] For there is little doubt Kirkland felt that success in her family's western adventure hinged upon the hope of civilization's inroads on the frontier; the village of Pinckney itself was a speculative venture upon the part of her and her husband Joseph.[11] So while the

success of *A New Home* may indeed have turned upon a readership to whom the text was "a sort of 'Emmigrant's Guide'" (*ANH,* p.1) and *caveat* to the westering individual, it seems no less certain that Kirkland's construct of a "far distant land" more west than Michigan embodies the counter-fantasy that her Pinckney readership had itself been borne *eastward* through expansion--- a fantasy which I would argue had been lodged in *A New Home* as a desire to remap the socio-economic spaces of the West itself, and in which the gaps between cultured and uncultured human activity are in reality the forms of desire for the stratified life of class-culture, marketplace economics, and the overcoming of frontier egalitarian attitudes.

For the well worn tropes of frontier independence and egalitarianism had meant something quite different to Kirkland, whose removal to the west had extracted her from a commodity world of goods and services, and had placed her in an economy where the lack of market embededness formed the foundations of western relations. Frontier exchange, to her mind, was established on the principle that "there are places where the 'almighty dollar' is almost powerless," having met with "its conqueror in the jealous pride of those whose services must be had in order to live at all" (*ANH*, p. 52). It was precisely this spirit of egalitarianism which Kirkland felt had altered labor relations and the acquisition of goods on the frontier, so that among persons who felt themselves to be equals, social relations had come to constitute more the figures of conciliation and concession rather than those of supply and demand which ruled the class-culture of the eastern seaboard. More succinctly, if frontier independence had (in the rhetoric of the day at least)

its foundations in the individual's desire to remain outside of the market for both goods and labor, one begins to understand more fully the equivocal nature of Kirkland's response to such values, since her purchasing power so outstripped her skills in domestic production that she soon found out that "[n]o settlers are so uncomfortable as those who, coming with abundant means as they suppose,... set out with a determination to live as they have been accustomed to live" (*ANH*, p.52).

Indeed, it is interesting to note that if Kirkland had recognized only after the publication of *A New Home* the need to conciliate her western neighbors, her narrator Mary Clavers had nonetheless construed it as a lamentable aspect of frontier economy as we see it in the text itself, where the difficulty of maintaining household help forms a central ideology of hardship, and a warning to those of her audience whom she expected to be "borne westward by the irresistible current of affairs" (*ANH* 50).[12] Obliged to employ a neighbor, Mrs. Jennings, "to 'chore round,' to borrow her own expression," (*ANH* 51) Clavers soon discovers that distinctions between wage-based relations and neighborly social relations are decidedly blurred in the west, and that hiring the woman out means nothing less than "the full enjoyment of her delectable society" (*ANH* 51). The breach of the employer- employee relationship is immediately contextualized here as unrefined, since Mrs. Jennings has the habit of "dipping with her own spoon or knife into every dish on the table," and drinking "the remains of the tea from the spout of the tea-pot, saying it 'tasted better so'" (*ANH* 51). Clavers abjures her eastern audience to "[l]et no one read with an incredulous shake of the head, but rather let my sketch of these peculiar habits of my neighbors be considered as a mere

beginning..." (*ANH* 51). And if, she further suggests, her readers think "it would be far better to starve than to eat under such circumstances," (*ANH* 51) let them consider that "[t]o have offered her outright a separate table... would have been to secure myself the unenviable privilege of doing my own 'chores'" (*ANH* 51-52). The desire to spatialize domestic relations, then, runs the risk of creating a scarcity of labor, even as its failure nearly devalues the social indice of maintaining help to begin with.

Yet material concerns consistently, if gallingly, outweigh social ones in *A New Home,* and while Clavers maintains that it "would be in vain to pretend that this state of society can ever be agreeable to those who have been accustomed to the more rational arrangements of the older world," (*ANH*, p. 52) she concedes that adaptability is an essential trait in securing one's needs, often by deprecating her own adherence to out of place customs. "Those who ... attempt to carry all sail set and to keep an unvarying course," she notes, "always abuse Michigan, and are abused in their turn" (*ANH*, p. 52). Such moments are diffused through transient claims to western identity, and Clavers suggests that those who make such "capital mistakes" wind up as failures, "revenging themselves by telling very hard stories about us nor'westers" (*ANH*, p. 52). Moreover, the concession to shared space is beyond the onus of failure once it is drawn into a progressive vision of frontier enlightenment; for if Clavers had come to understand that " it is unwise to stem directly the current of society," *ANH*, p. 52), she could likewise assert that more indirectly, "the silent influence of example is daily effecting much toward reformation. (*ANH*, p. 53). Lodging the discourse of social reform within the cultivation of virtue at the table, dining itself becomes the occasion for a treatise on

the reawakening of status and deference in social relations, as Clavers notes that "[n]eatness, propriety, and that delicate forbearance of the least encroachment upon the rights... of others has only to be seen and understood to be admired and imitated" (*ANH*, p. 53). Restricting the rhetoric of egalitarianism to the sphere of inalienable human rights, this idyllic vision of the table nonetheless asserts the concept of status as the gift of culture, and assumes that when equals in the eyes of God meet in worldly relations, they naturally recognize the social distinctions that exist between them.[13]

Though given over to the syntax of optimism and futurity, this reverie remains essentially bound up with the rhetoric of early national republicanism, primarily as a prepartisan fantasy in which social order is guaranteed through the concepts of individual virtue, deference to the rights of others, and the consistent identification of self-interest with the common good.[14] So that while Mrs. Clavers could confess that her own concessions were strategies of "*wearing round*, which give me the opportunity to live very much after my own fashion" (*ANH*, p. 52), she was likewise able to suggest that the virtue of such action resides as well in its model of civic improvement, the patient instructor of manners having discovered "ere long that [her] neighbors have taste enough to love what is so charming, even though they see it exemplified by one who sits *all day* in a carpeted parlor" (*ANH*, p. 53).

Given the historical context of Clavers' claims, however, her paeans to republican values acquire something of a weak cadence in a period which Joel H. Silbey has characterized as marking a shift from "the colonial revolutionary one of classical republicanism, with its deferential

process, limited democracy, and antiparty ethos, to the new world of egalitarian partisanship, with organized... political warfare shaped by powerful national parties."[15] Nineteenth-century Americans were forced to confront an alternative vision of the body politic which Michael Holt argues "seemed volatile, fluid, and constantly open to the possibility of change," and which had discarded the republican idea of a fragile *polis* secured through civic virtue for one in which the check and balance of competing interests was a more natural model of political process.[16] For Clavers, this had suggested that the contested ground of partisan politics might itself be the source of strife in the domestic economy, especially so since sectional interests had to her mind secured the frontier for Jacksonian Democracy.[17] The platform which had asserted independence through freely held property and market aloofness had so disrupted labor markets she could complain that political involvement was itself the hallmark of low productivity. Exasperated by the difficulties she encountered in completing the construction of a framed dwelling for her family, she states she had more than once visited the site only to find herself "in complete solitude, every soul having gone off to meeting or town election" (*ANH*, p. 48).

Clavers infers that no one's interests are well-served in this distraction, suggesting to a workman that it "must be a great sacrifice... to spend two days in going to election" (*ANH*, p. 48). She is made to understand, however, that the help is as well-schooled as she in the discourse of republican values, and his rejoinder that "ought not a man do his duty to his country" (*ANH*, p.48) effectively displaces the wage market as the ground of interest in the common good, and asserts that within

the exercise of "the noblest privilege of a 'freeman'" (*ANH*, p. 48) lies the true capacity of the individual to "forward your business and his own" (*ANH*, p. 48). "This was unanswerable, of course," (*ANH*, p. 48) Clavers notes, for the claims to republican virtues are apparently open as well to Jacksonian Democrats operating in the world of partisan politics and majority rule.[18] What remains is for Clavers to make the accusation that reticence in the labor market constructs a political economy which is incompatible with domestic economy, since the privileged claims to duty within the sphere of family are the most likely to compensate for the assertion that a broader civic responsibility exists solely within an arena of male values. Clavers insists that male wages and productivity bear directly upon the value of domestic comfort, and it is "poor Mrs. Fenwick, whose tattered gown would have been handsomely renewed by those two day's wages" (*ANH*, p. 48)who pays the price of civic virtue, and whose torn dress reveals the moral breach at the core of a house "not built in a month, nor two, nor yet in three" (*ANH*, p. 48).

It is unclear the extent to which the above model of exchange evokes a market intrusion into women's productivity in the home, though it clearly undermines the purchasing power of frontier wives who appear to be at least to some degree dependent upon marketed goods at this level of the domestic economy. Such a distinction, however, re-examines only qualitatively the nature of women's labor on the frontier, since as Nancy Grey Osterud has suggested, the distribution of "labor 'rationally' according to its productivity ... and the relative advantages of producing or of purchasing their subsistence" was simply one manner in which early American farm families thought about the value of their labor in relation

to time.[19] Yet Clavers implies that productivity geared toward the rationale of the market remained sporadic at best, a neighbor having informed her that her search for domestic help is limited to the single prospect of "a young lady here yesterday that was saying she did n't know but she'd live out a spell till she'd bought her a new dress" (*ANH*, p. 38). This is "never as a regular calling," (*ANH*, p. 39) however, and the irrationality of such a state of affairs is lodged in her assertion that "I have ... seen the interior of many a wretched dwelling,... nothing in it but a bed, a chest, and a table... and potatoes the only fare; yet never yet saw I one where the daughter was willing to own herself obliged to live out at service" (*ANH*, p. 39).

The issue continues to rest with Clavers' own productive capacities, however, since the goods and services upon which she relied she felt could be secured only within the developed forms of a market economy. "The division of labor is almost unknown," (*ANH*, p. 72) she complained, "so in the state of society which I am attempting to describe, each woman is, at times at least, her own cook, chamber-maid and waiter; nurse, seamstress and school-ma'am; not to mention various occasional callings to any one of which she must be able to turn her hand at a minute's notice" (*ANH*, pp. 72-73). Indeed, to the woman who "fell climax!-- knows nothing at all of soap-making," (*ANH*, p. 53) frontier inhabitants (male and female alike) display an almost hyperbolic level of practical skills which, in the west, can turn the tables of social order:

> Is one of your guests dependent upon a barber? Mr. Jenkins can shave. Doesyour husband get *too* shaggy? Mr. Jenkins cuts hair. Does he demolish his boot upon a *grub?* Mr. Jenkins is great at a

> *rifacciamanto....* Does your bellows get the asthma? Mr Jenkins is a famous Francis Flute (*ANH* 80).

None of these skills, however, amount to "a regular calling," either, though Clavers was able to register her optimism in a vision of specialization as a budding marketplace. "[H]alf-fledged as it is, " the village soon boasts of "the blacksmith, the Cooper, the chair maker, the collar maker... the mantua-maker for your dresses,... the plain seamstress, plain enough sometimes," (*ANH*, p. 80) and others who traffic in particular goods and services.

More acute, perhaps, is the sense that women regularly find themselves dependent within the more narrow sphere of domestic economy, where, for example, a pragmatic gift for managing space in confined quarters, and the understanding that "'[w]here there ain't no room for a thing, why, there ain't'" (*ANH*, p. 45) allows the temporary Angeline to "cut the Gordian knot" (*ANH*, p. 45) of struggling to reconcile too little space with too many goods. "I was forced to confess," Clavers states, "that she knew much better than me how to make a log-house comfortable" (*ANH*, p. 44). However, the recognition by more seasoned women that interdependence in the sphere of the domestic economy is essentially at the core of their overall relations tends to deflect the sort of anxiety which Clavers lodges within this admission. Indeed, there are moments when the nature of social discourse tends to subsume the business of frontier economy in *A New Home,* revealing the manner in which socio-economic relations constitute the very fabric of life in Montacute. Taking afternoon tea with neighbors, for example, Clavers records the following conversation:

> "Do you have any butter now?" "When are you going to raise your barn? "Is your man going to kill this week? "I ha'nt seen a bit of meat these six weeks." "Was you to meetin' last Sabbath?" "Has Miss White got any wool to sell?" "Do tell if you've been to Detroit!" "Are you out o' candles?" "Well I *should* think Sarah Teals wanted a new gown!" "I hope we shall have milk in a week or two..." (*ANH*, p.137)

Social gossip blends readily here with images illuminating a network of labor, barter, lending, and moneyed exchange, and a true picture of frontier economy emerges as the circulation of a small amount of surplus goods and labor, distributed in such a way as to form the occasions in which women look *and* speak to the comforts of their neighbors' homes on a regular basis.[20] And while Clavers' own discomfort may often be registered through the inability to deal with excess in cramped quarters, "let it be known," she nonetheless declares, "that in a state of society like ours, the bare necessaries of life are subjects of sufficient interest for a good deal of conversation" (*ANH*, p. 137).

Yet one looks in vain for Clavers' own voice within the undifferentiated flow of linguistic and economic production that constitutes this passage. And again, as a woman who "knows nothing at all of soap-making," (*ANH*, p. 53) the questions which surround her ability to fully participate in such an economy continue to inform the sense of alienation in *A New Home* which cannot be overcome even in its most genial moments. If she had come to understand that "however we may justify certain exclusive habits in populous places, they are... confessedly ridiculous in the wilderness," (*ANH*, p. 65) she discerned nonetheless that exclusivity was Janus-faced on the frontier, and that to belong fully to Montacute society, she must "at all sacrifice... live down the impression that I felt *above* my neighbors" (*ANH*, p. 65). Whether or not Clavers

views such concessions with distaste is little to the point; for if the above passage reveals that living equally with one's neighbors was to be immediately bound up with them in socio-economic relations, the issue remains the capacity, and not the will for reciprocity, and Clavers can only imply that she has social sacrifices to make in return for materials and services. Registering her own determination to participate more fully in the social /material market-place, she chides herself "[i]f I treat Mrs. Timson with neglect today can I with any face borrow her broom tomorrow? And what would become of me, if in revenge for my declining her invitation to tea this afternoon, she should decline coming to do my washing on Monday?" (*ANH* 65). There seems nothing particularly disingenuous about this sentiment, and the idea that social gifts bear a fair market value is, after all, only consistent with the value inherent in disseminating culture at the table, and the process of social mediation was equally served for Clavers in her observation "that 'levelling upwards' is much more congenial to 'human natur,' than levelling downwards" (*ANH*, p. 185).

It remains to be seen, however, whether or not her neighbors believe that such gestures do much in the way of effacing class distinctions based on obvious differences in material wealth. To be sure, there is little in *A New Home* which manages to deflect these class realities so long as the structures of the sketch continue to foreground the physical aspect of the home itself as the ideology of social stratification.[21] Visitation carried on in the dwellings of Montacute families is the central zone of class confrontation in the text, and though Clavers states that she "took especial care to be impartial in my own visiting habits," (*ANH*, p. 65)

she nevertheless implies that such social dictates can be trying, especially when "a neighbor happen[s] in after breakfast for a three hours' call," (*ANH*, p. 65) or when "the knitting needles" which "might peep out beneath the shawl" (*ANH*, p. 65) reveal the manner in which one's own time and space can be readily appropriated through the codes of congeniality.

Visitation is never purely social in *A New Home*, however, and frontier life often brings with it hardships which, as Clavers noted, constitute the numerous occasions which "may throw you entirely upon the kindness of your humblest neighbor" (*ANH*, p. 65). The ague, in particular, is a constant bother in the west, and its habit of striking across gender and class distinctions often forms a leveling backdrop against which less "natural" social stratifications can be observed, since, as Mrs. Titmouse, a Montacute neighbor observes, "if you've got to have it, why you can't get clear of it," (*ANH*, p. 117) and Clavers herself suggests that "[m]oonlight and the ague are... the same every where," for "did not great Cæsar shake 'when the fit was on him?'" (*ANH*, p. 115). Having experienced, firsthand, then, the difficulties of managing on one such occasion, when the neighbors had "showed but little sympathy," and having "imbibed the idea that we held ourselves above them...chose to take it for granted, that we did not need their aid," (*ANH*, p. 61) she sets out charitably with a friend in Chapter XXX to visit the ailing Titmouse family, the "fit" being on them.

The opening statements in this chapter form an interesting moment of self-reflexive discourse dealing with both the figures which come under

scrutiny in *A New Home*, and the manner in which those figures can be represented in relation to traditional aesthetic models:

> How can one hope to find any thing of interest about such common-place people? Where is the aristocratic distinction which makes the kind visit of the great lady at the sick-bed of suffering indigence so great a favour, that all the inmates of the cottage behave picturesquely out of gratitude--- form themselves into *tableaux*, and make speeches worth recording? Here are neither great ladies nor humble cottagers. I cannot bring to my aid either the exquisite boudoir of the one class, with its captivating *bijouterie*... nor yet the cot of the other more simple but not less elegant... and inhabited by goodness, grace and beauty. (*ANH*, p. 115)

Clavers here makes play of literary genres which promote a sort of rustic Utopia and the assurance of class harmony through an idealization of estate-based production, and which Raymond Williams has argued "have been commonly read as describing an actual rural economy: an existing social base for the perpetual peace and innocence of the neo-pastoral dream."[22] She maintains no such allusions, however; "[t]hese materials are denied me," she states, and "yet I must try to describe something of Michigan cottage life" (*ANH*, p. 115). And if, she implies, she runs the risk of repeating these transgressions through the necessity to use "such delicate periphrasis as may best veil the true homeliness of my subject," (*ANH* 115) the fault lies in "the delicate organization of 'ears polite,'" (*ANH*, p. 114) and not in her own understanding of rural economy.

Yet what follows is nothing short of the appropriation of narrative intent by one of those figures of cottage life, Mrs. Titmouse, "whose untamable tongue was too much even for the ague" (*ANH*, p. 116).[23] Greeting the well-wishers at the door of "an ordinary log house... quite old and dilapidated," (*ANH*, p. 116) she registers her own view of the situation, and cries "Bless my soul! if I ever thought to be catch'd in sitch

a condition, and by sich grand ladies too!" (*ANH* 116). The true nature of the condition is equivocal, however, and it is unclear whether she speaks of the ague, or of the social trappings of her home, so distinct from what she perceives to be Clavers' own sense of comfort she confesses to her that she "can't show ye no civility" (*ANH*, p. 116). With an effusion of rustic politeness nonetheless, she tells the ladies "[d]o sit down... if you *can* sit upon a chest" (*ANH*, p. 116). Such talk goes on, Clavers suggests, till the "poor soul ... was fairly out of breath," Mrs. Titmouse complaining that Mr. Titmouse "ha' n't no ambition," and so clearly marks out the deficiencies of their state that the reader is forced to agree when she says "we live jist like the pigs" (*ANH*, p. 116).

Fearing, perhaps, that Mrs. Titmouse's "untamable tongue" is preparing to quite undo the gesture of social mediation implied in this act of kindness, Clavers and her friend Mrs. Rivers attempt to "out-talk her with ...assurances that we could accommodate ourselves very well" (*ANH*, p. 116). Finding a gap in the unabated flow of discourse, "tell us," interjects Mrs. Rivers, "what we can do for you" (*ANH*, p. 116). This hardly has the desired effect, however, and in the salvo Mrs. Titmouse launches in response to this gesture, there is inscribed not only the reality of class structure as she sees it, but also a re-evaluation of the comparative virtues of one class against another where practical skills command the business of day-to-day affairs: "Do for me! Oh, massy! Oh, nothing, I thank ye. There a' n't nothing that ladies like you can do for me" (*ANH*, p. 116).

Her husband, however, is a practical man, telling his wife roundly "[y]ou know we ha' n't tasted a mouthful since morning" (*ANH*, p. 116).

"[U]nder a thousand compliments" (*ANH*, p. 116) of Mrs. Titmouse, then, the ladies are prevailed upon to prepare a bit of food for the family. The ensuing scene fairly bears out Mrs. Titmouse's judgement, and the two friends barely manage to set a loaf of bread to bake, and are unable to procure much more water than will "half-fill the tea-kettle" (*ANH*, p. 117). This much accomplished nonetheless, and having "done what little we could," the two prepare to make their leave, "glad to escape [the] overwhelming protestations of eternal gratitude," (*ANH*, p. 117) though Mr. Titmouse suggests that "the cow had not been milked since the evening before" (*ANH*, 117). "Here was a dilemma!" Clavers states, "How we regretted our defective education, which prevented our rendering so simple yet so necessary a service to the sick poor" (*ANH*, pp. 117-18). Still, the women decide to "set [them]selves resolutely at work to ascertain [their] powers in the milking line" (*ANH*, p. 118). But as "the cow gained the victory, as might have been expected" and the two "were fain to retreat into the house," (*ANH*, p. 118) the scene ends with the Mrs. Clavers and Rivers supporting the ailing Mrs. Titmouse while she does the milking, and "having been partially successful in this," the reader is left to ruminate once more upon this woman's "untamable tongue," as the two escape to the woods hearing "Lord bless ye! I'm sure I'm under an everlastin' compliment to ye; I wish I know'd how I could pay ye. Such ladies to be a waitin' on the likes of me; I'm sure I never see nothing like it, &c. &c." (*ANH*, p. 118).

What had begun, then, as an attempt to redefine literary representation in *A New Home* as distinct from the ideological motives of rustic fantasies, ends having done so only at the cost of redefining the

nature of the women's gesture itself. "In cities," Clavers will suggest at the end of *A New Home*, "we bestow charity--- in the country we can only exchange kind offices, nominally at least" (*ANH*, p. 184). The cow, we are led to believe, "had small respect for timid and useless town ladies," (*ANH*, p. 118) but it is the persistence of Mrs. Titmouse's declarations that she has fallen under "everlastin' compliment," and Clavers' own commitment to regard "as speeches worth recording" (*ANH*, p. 115) the untamed discourse of her neighbors, which cuts both ways. She had sought, in other words, to make real a vision of social harmony which she viewed as specious in traditional pastoral images, only to discover that such a vision did not rest upon her impressions alone. Neither kindness, nor charity, it seems, can truly begin at home, so long as those who must be receptive to these acts continue accept them after their own fashion, and in their own social space.

If Clavers had suggested that the types of aesthetic conventions which had cottagers behaving "picturesquely out of gratitude" (*ANH*, p. 115) might ultimately fashion a ground of social misreading, it could be argued as well that she had hoped that "composing" her work within the popularized genre of literary sketches might, after all, complete the "search for the 'union' of 'variety' and 'harmony,' for the resolution of potentially antagonistic forces" inhabiting the social landscape. [24] Yet there are other moments in the text where the status of representation threatens to break down altogether, and if Clavers was unable to harmonize the social stratification inherent in her relations with the Titmouse household, she suggests elsewhere that there are certain elements of Montacute society whose condition so differs from her own

that her claims to verity lay only in that she had "attempted to give some idea of the appearance of things" (*ANH*, p. 109), but could not "pretend to paint the confusion" (*ANH*, p. 109) at the center of heightened moments of frontier poverty. The Newlands, for instance, whose "shanty built against a sloping bank, with a fire-place dug in the hill-side, and a hole pierced through the turf by way of chimney (*ANH*, p. 108) is another example of moral architecture in *A New Home,* belong, she states, to "one class of settlers whose condition has always been inexplicable to me" (*ANH*, p. 107).

Clavers had, at the outset of *A New Home*, given the warning to "any fashionable reader who may have taken up my book, that I intend to be 'decidedly low,'" (*ANH*, p. 4) and to all appearances, this depiction of a family whom, as she states, "had burrowed" (*ANH*, p. 108) rather than lodged in their home examines to the very limits this end of Montacute society. Yet where the rhetoric of intentions in such a caveat had implied that Clavers herself had felt she understood how far she might portray questionable aspects of Montacute life, her discussion of this hard-pressed family in Chapter XXIX appears compromised throughout by a good deal of indecision, and by what must appear to the twentieth-century reader at least, a troublesome lack of understanding in her own relations to a family who had, after all, only "fallen under [her] notice" because "[t]he father and his... sons had performed a good deal of hard work in our service, and the females... had been employed on many occasions when 'help' was scarce" (*ANH*, p. 108). For while Clavers manages adeptly the "decidedly low" representation of this family's material conditions, they seem to be nonetheless *misrepresented* insofar as she cannot fully articulate the

economic conditions which go into creating them. These moments in which Clavers finds certain conditions of frontier life "inexplicable" are, however, as indicative of *A New Home* as its most honest and lucid revelations, and the reader seeking out the realism of the text needs only to look past the limits of the narrative to see there the markers of a much broader set of relations governing Montacute life.

"They seem to work hard," Clavers states of that class of settlers characteristic of the Newlands, "[t]hey perform the severe labor which is shunned by their neighbors" (*ANH*, p. 107). This relegation to hard physical labor, and the instability both financial and domestic in circumstances where "even mamma, the center of dignity is fain to go out washing by the day," (*ANH*, p. 108) constitutes indeed something of a re-examination of the market for labor which had been so problematic for Clavers earlier on in the text. And though she is careful to instil within these individuals the same "peculiar air of dignity and 'dont care,'" (*ANH*, p. 107) which is characteristic of Jacksonian principles, it is clear that the Newlands represent something of a new class of wage-dependent agricultural laborers, indicative, perhaps of their terrestrial surname. For it may well be that Clavers had misjudged, somewhat, that "current of affairs" which had borne so many westward in the 1830's, attributing most of the emigration to government policies which sought to provide cheap land in hopes of broadening production, as well as creating a burgeoning market for consumer goods produced in the northeastern industrial centers. However, conditions prior and leading up to the Panic of 1837 had displaced much of the eastern populous, constituting both farmers and an urban class of skilled craftsmen, artisans, and laborers.[25]

Clavers herself had taken note of this class of emigrants precisely in relation to the Michigan land boom, and suggested their fates at the hands of dishonest speculators:

> [I]t must have required some nerve to carry the matter through; to stand by, while the poor artizan, (*sic*) the journeyman mechanic, the stranger who had brought his little all to buy government land to bring up his young family upon, staked their poor means on strips of land which were at that moment under a foot of water. (*ANH*, p. 31).

That Clavers may have failed to see in this image conditions which had uprooted, rather than beckoned this element of westward migration suggests that she had difficulty in broadening the perspective of *A New Home* beyond a discourse of opportunity, and future hopes.

It remains to be seen whether or not the privilege which the future wields over the past is consistent along gender lines, yet the fact is that the Newlands are the one family Clavers takes up at length who do *not* have a story to tell about their past, nor of their removal to Michigan, and so there is no other marker of their status readily available to the reader than their necessity to utilize labor normally expended in agricultural production to "go to service," or to "'chore round' for every body and any body" (*ANH*, pp. 107-08). Discernable, however, in their condition of being reduced to the purchase of the "coarsest food," the need to "ask for an old coat or a pair of shoes," and most significantly, the "increasing dearth of cows, pigs and chickens" (*ANH* 107) populating their yard is the reality that one needs, even in the west, a certain amount of capital to produce any surplus beyond that which meets the requirements necessary to reproduce one's own capacity for work. And indeed, to Clavers' own sensibilities, it appears that the Newlands get along by "apparently

denying themselves... everything beyond the absolute necessaries of life" (*ANH*, p. 107).

What is striking in this assertion, however, is that it is raised not in terms of the means of living which short wages will purchase on the frontier, but rather that these "apparently" ascetic principles are contextualized within an ideology of thrift and economy, so that what is indeed "inexplicable" to Clavers is the paradox that "instead of increasing their means by these penurious habits, they grow poorer every day" (*ANH*, p. 107). Unable to rationalize the scarcity of the most basic goods within the Newland's apparent willingness to embrace all out the ethic of hard work and domestic economy, Clavers confronts directly the mythos of opportunity in the west, which had indeed rested upon the ideology of cheap land and small, productive farms, yet which had done little to account for the capital investments necessary to undergo such a venture.[26] An unannounced visit to the Newlands, however, celebrating a removal to a "log-house of the ordinary size... though it had neither door nor window, nor chimney nor hearth," (*ANH*, p. 108) offers the occasion to deflect this contradiction, and the remainder of the chapter re-situates the Newlands and their "class but too numerous in Michigan" (*ANH*, p. 111) within what Charles Sellers has declared the "dominant mythology" of the era, "endlessly proclaimed by the successful, that opportunity was rife and that success or failure turned on effort and character alone."[27]

"[W]e were enjoying the fine sleighing," Clavers recounts, when it "so happened that we found ourselves in the neighborhood of Mr. Newland's clearing" (*ANH*, p. 108). There is, of course, a glaring irony in the implication that leisure activities conduct the Clavers' to a log-house

where, as she notes, "a quilt is but a poor substitute for a window during a Michigan November" (*ANH*, p. 108). However, what the occasion furnishes more readily is the device by which the paradox of appearance and reality inherent in the Newland's circumstances can be undone, "so noiseless was our approach, guiltless of bells, that no one seemed aware of our coming" (*ANH*, p. 108). Expecting, perhaps, a show of "penurious habits" which would, after all, only continue the "inexplicability" of how such hard work can produce so little by way of comfort, Clavers instead discovers "Mrs Newland... evidently in high holiday trim," (*ANH*, p. 108) and is given to understand by Mr. Newland, that "the girls thought they must have a kind of a house-warmin' like" (*ANH*, p. 109).

The scene of this festivity seems harmless enough. However, the moral implications of this "accidental discovery" (*ANH*, p. 110) immediately set to work rendering explicable the destitution of a family whom Clavers says "had always appeared before us in such different characters" (*ANH*, p. 109). The "more striking part of the picture," she states, is the dress of the Newland children, the boys "flaming in green stocks... while the cut of their long... coats showed that whoever they might once have fitted, they were now exceedingly out of place," and the girls, whose "tawdry, dirty finery" left them "wearing any look but that of the modest country maiden" (*ANH* , p. 109). There is, as well, an incident in which an ox-sled, "loaded with young people of both sexes" is overturned "into the deep snow, by a 'mistake on purpose' of the driver" (*ANH*, p. 109). And though one suspects that the nineteenth-century reader was a more discerning judge, and would have found fault enough in this alone to have understood the implications of the scene, the most

damning object of scrutiny is apparently the dinner table, upon which "a cloth seemed to cover a variety of refreshments," and in front of which "stood a tin pail, nearly full of a liquid whose odour was but too discernible" (*ANH*, p. 109).

Clavers espouses temperance values freely throughout the text; yet it needs to be stated that the image of alcohol, when it appears in *A New Home*, is firmly rooted within an ideology of productivity and consumption, and the reader discerns readily enough the inverse relation of the one to the other. Hence, when a barn raising in chapter XII results in a smashed thumb, and a blow to the head, "[a] jug of whiskey was pointed out... as the true cause of the matter" (*ANH*, p. 41). Clavers then goes on to state that "[t]he many raisings which have been accomplished at Montacute, without that ruinous ally, strong drink... have been free from accidents of any sort" (*ANH*, p. 41). An obviously abused woman with whom Clavers boards in Chapter I laments her "change of lot-- from a well-stored and comfortable home.. to this wretched den in the wilderness," a fate which is "owing to the horrible drunkenness of the master of the house" (*ANH*, p. 7). For a destitute band of Native Americans in Chapter IX, whiskey is "the baleful luxury which performs among their race the work of fire, famine, and pestilence" (*ANH*, p. 29). And the neighbor woman Mrs. Ketchum attempts, "like many poor souls... to hide her husband's real difficulty" by suggesting an *excess* proclivity for work, exclaiming "He was *so* tired!" (*ANH*, p.37) in response to his having fallen asleep at the dinner table.

All of these images finally come home to roost, however, in Clavers' suspicion that a vision of surplus agricultural production could itself be co-

opted and rendered immoral by excessive and wasteful consumption, "turning our fields of golden grain into 'fire water'-- a branch of business in which Michigan is fast improving" (*ANH*, p. 7). Notwithstanding the probable realities of Clavers' assertions, then, alcohol becomes, in *A New Home*, the nexus of an economic ideology which seeks to link material production to moral uprightness in an allegory of frontier opportunism. And as for the Newlands, the reader is given to understand that "penurious habits" are merely the veil disguising the specter of inconspicuous consumption, the root cause of "the utter discrepancy between [their] complaints, and the appearances we had witnessed (*ANH*, p. 110).

Finally, Clavers works this model of immoral production into the expenditure of the body itself, for she had earlier remarked that one of the Newland daughters, Amelia, "was arrayed in far better taste than her sisters', and by contrast with the place and circumstances....[h]er dress was of rich silk, made in the extreme mode, and set off by elegant jewelry.... most elaborate pendants of wrought gold hung almost to her shoulders; and above her glittering basilisk eyes, was a gold chain with a handsome clasp of cut coral" (*ANH*, p. 109). The reason she can't wear this attire in public, one suspects, has to do with the fact that she "had been lately at a hotel in a large village at some distance, and had returned... not improved either in manners or reputation" (*ANH*, p. 109). Prostitution, of course, would have formulated for Clavers' readers that "decidedly low" end of moral indiscretion. However, to consider that there is inscribed here as well the iconology of a perverse work ethic is, after all, only to efface the distinction between the moral implications of

sleeping around, and Clavers' own judgement upon wage laborers who "go to service and 'chore round' for any body and every body" (*ANH*, pp. 107-08). That Clavers suggests here the wage-contract ought surely to be as monogamous as sexual relations within the marriage contract is something to be considered, recalling her statement to Mrs. Ketchum, that "I wish to get a girl who will remain with me; I should not like to change often" (*ANH*, p. 38).[28] More to the point, however, is the implication that immoral productivity will yield only those sorts of returns which fall under "the 'slow unmoving finger' of public scorn," (*ANH*, p. 111) and which, most significantly, cannot be displayed as the registers of Montacute's moral and material growth. For not only is Amelia's dress hidden from public gaze, but the reader who can bridge the temporality of her hotel stay with the horrific scene which ensues will understand that the remainder of this critique concerns the costs of illicit reproductivity, and Clavers goes on to exclaim "How was I surprised some two months after at being called out of bed by a most urgent message... that Amelia... was dying!" (*ANH*, p. 110).

The apparent victim of a botched abortion, Amelia has expired before Clavers can arrive, and as a result the incident is sufficiently censured from gentler eyes by the narrator's inability to speak as a witness, as well as by the reticence of everyone present, one who "could give no account of her condition," and another, who, when "asked for particulars... shook her head and declined answering" (*ANH*, p. 110).[29] However, the girl's corpse, "swollen and discoloured, and already so changed in appearance that I should not have recognized it earlier," (*ANH*, p. 110) remains as a text through which Clavers can now register a

more social response to this "class of settlers," stating that "[r]umour was busy as usual; and I have been assured by those who ought to have warrant for their assertions, that this was but one fatal instance out of many *cases*" (*ANH*, pp. 110-11). "[T]he whole appearance of this sad wreck was quite different from that of any corpse I had ever viewed before," (*ANH*, p. 110) she states, though it is precisely through the chapter's sustained attempts to make appearance and reality comply that she manages to suggest that this wretched "body public" has been rejoined with the "basilisk eyes," and "dress...made in the extreme mode" of the "body private." The reader is left to ruminate, then, on the value of putting aside the wages of sin for those of monogamy in more healthy labor relations.[30]

Earlier, in Chapter XVII, Clavers had called on her friend Mrs. Rivers to attend a wedding, and, having "found her in her ordinary neat home-dress, had "induce[d] her to exchange it for a gay chally with appropriate ornaments" (*ANH*, p. 65). "It really seems ridiculous," was the reply, "to *dress* for such a place! and besides... I am afraid we shall be suspected of a desire to outshine" (*ANH*, p. 66). Speaking for the natives, Clavers' response is to alert her to "that other and far more dangerous suspicion of undervaluing our rustic neighbors," (*ANH*, p. 66) and one sees here how readily the ideology of dress supplies the indices of social relationships. There is irony, however, in the fact that Mrs. Rivers is prevailed upon to wear "any look but that of the modest country maiden," (*ANH*, p. 109) and what emerges here is indeed a recontextualization of the relations between domestic economy, performed in "neat home-dress," and the social markers of that economy, the "gay challys" and

"appropriate ornaments" earned through ascetic housekeeping and "penurious habits." It is, then, Amelia's failure to sufficiently link her dress "in the extreme mode" with a more appropriate sphere of productivity, that constitutes the core of her reputation. And while she may not have "undervalued" her neighbors so strongly as Clavers would suggest, she had certainly understood for herself the implications of "outshining" them.

"[W]ondering no longer," then, "at the destitution of the Newlands," (*ANH*, p.109) Clavers proceeds to chart their removal from the textual geography of Montacute, "driving off with their own, as many of their neighbors' cattle and hogs as they could persuade to accompany them" (*ANH*, p. 111). It will be remembered, however, that this migration is also a subtle remapping of Montacute itself, for it is now an extended version of the frontier which is "rapidly drafting off those whom we shall regret as little as the Newlands," (*ANH*, p. 111) and Clavers is able to depict through their absence a more stable vision of village life. "I trust we have few such neighbors left," she states, "Texas and the Canada war have done much for us in this way" (*ANH*, p. 111). The Newlands are given the task of carting off to the "wide west" the specter of their productivity as well, in an image hard to overestimate in its demoralization of a class which Clavers states "is a vicious and degraded one" (*ANH*, p. 111). Transcribing the Newland's "[m]any requests for cast articles" and "such comforts as our house afforded" (*ANH*, p.108) into a perverse model of idle subsistence, the Newlands' departure is accompanied by a "train of fierce dogs which have not only shown ample sagacity in getting their own

living, but... assisted in supporting the family by their habits of nightly prowling" (*ANH*, p. 111).

The passage raises the issue once again of regression as the model of westward migration; St. John de Crèvecoeur, for example, had suggested a quite similar view of life in these "last inhabited districts," where "men appear to be no better than carnivorous animals of a superior rank," and where, "wholly left on their native tempers.... many families exhibit the most hideous parts of our society." [31] He had as well, however, utilized the paradigm of westward migration to chart an ideology of improvement, not unlike that of Clavers, suggesting that in the wake of those whom "vice and the law" had driven west, would follow a more industrious and genteel class of settlers who "will finish their improvements, convert the loghouse into a convenient habitation, and...will change in a few years that hitherto barbarous country into a fine fertile, well regulated district.[32]

Yet while Clavers had, in *A New Home*, continued to depict the slow pace of improvement in Montacute within the idiom of a stabilizing economy, until she could boast that the town, "half-fledged as it is, affords facilities that one could scarce expect," (*ANH*, p. 80) she had little recourse to more privatized images such as Crèvecoeur's, which relied upon ownership and an agrarian model of wide ranging cultivation as the index of a moral society. For while the latter had been able to remark upon the forty-year's worth of improvements on a piece of property grounded in inheritance from his father, the Michigan land boom of the 1830's had so destabilized long-term ownership that Clavers was conversely able to portray within the image of banknotes backed by property a ludicrous

vision of land as a moveable asset, set upon by speculators who, more often than not, took the profits from shady business deals to furnish "splendid metropolitan residences," where "burr-oaks were turned into marble tables, tall tamaracks into draperied bedsteads, lakes into looking glasses, and huge expanses of wet marsh into velvet couches, and carpets from the looms of Agra and of Ind" (*ANH*, p. 121).[33] "Real-estate pledged," she stated, "the notes were better than gold or silver," because they were lighter in the pocket" (*ANH*, p. 122).

Moreover, it seems clear that in Clavers' inability to call upon the vision of a "fine fertile, well regulated district," there is implied the limitation of the sphere of representation available to women, whose sense of stability and productivity was confined largely to the domestic sphere. She had indeed suggested that an agrarian life *anywhere* offered a model of enduring stability for the male who, in removal, nonetheless "gazes on the same book of nature which he has read from infancy," while woman, whose relocation means adaptation "has been looking in vain for the reflection of any of the cherished features of her own dear fire-side" (*ANH*, p. 146). There is as well a temporalization of these separate spheres which , as Clavers states, allows the male to look to the future, "strong in heart and full of self-gratulation on the favourable change in his lot," while a woman can only register her new environment in terms of its distance from past comforts, causing "the time-honoured cupboard" to be "meagerly represented by a few oak-boards lying on pegs and called shelves" (*ANH*, p. 146).

It would appear, then ,that within the ideological representation which marks westward migration as both material and moral regression,

and with which Clavers had sent the Newlands on their way, there is little to alleviate the possibility that the domestic sphere might simply continue to be registered in lapserian terminology. For it had been Crèvecoeur's strategy to locate what Raymond Williams has called an "ideology of improvement" within a vision of increased agricultural production, while Clavers could only note within her limited sphere a consumer's response to such a view, arguing that "[t]he conviction of good accruing on a large scale does not prevent the wearing sense of minor deprivations" (*ANH*, p. 147) at home.[34] "Minor deprivations," to be sure, retranslated in the extreme the gendered view of Crevecoeur's "convenient log-cabin" into a "wretched den in the wilderness," (*ANH*, p. 7) and Michigan architecture itself made of the woman's "own dear fire-side" a "great open hearth" (*ANH*, p. 13) whose suffocating and "inextinguishable fire," Clavers noted, was something she "had never happened to see alluded to in any of the elegant sketches of western life which had fallen under [her] notice" (*ANH*, pp. 48-49).

It is, however, the gnawing concern that one's own character might suffer regress in such conditions which finally structures the representation of material life on the frontier, and in Clavers' own response to western egalitarianism, that "one cannot help but observing that 'levelling upwards' is much more congenial to 'human natur' than levelling downwards," (*ANH*, p. 185) there are discernable the forces at work which continually run against the grain of that assertion. And in response to such forces, she implied that one can only "*simplify* the matter very much," which meant for her that "[t]he children had dirty faces, and aprons which would have affected their total exclusion from genteel society" and that she would have to be content "to encourage the closest

intimacy between them and the calves and chickens, in order to gain some peace within doors" (*ANH*, p. 40). While Clavers constructs an inter-species intimacy here more benign than that which she applied to the Newlands, the similarities are hard to gloss, and where such "intimacy," moreover, becomes the marker of exclusivity from "genteel society," Clavers risks, if only for the moment, a re-alignment of her own narrative voice and its relation to her eastern audience.

Indeed, one might suggest that a narrative of seduction so popular in sentimental fiction finds its place, in *A New Home,* within the ongoing struggle to avoid an irrevocable fall into intimacy with western society.[35] Clavers herself noted near the close of the text that in certain points touching refined society, she had "already been adventurous, far beyond the bounds of prudence, " and stated as well that "in acknowledging even a leaning toward the 'vulgar' side, I place myself forever beyond its pale" (*ANH*, p. 186). The ideology of communal property in the west, for example had suggested how far such intimacy could extend, and Clavers was continually bothered by the fact that "[n]ot only are all kitchen utensils as much your neighbors as your own, but bedsteads, beds, blankets, sheets, travel from house to house, a pleasant and effectual mode of securing the perpetuity of certain...peculiarities of the skin" (*ANH*, p. 67). Progressing from the imagery of the transfer of infectious ailments, Clavers then proceeds to extend this model of familiarity to the maternal, recounting an instance in which one nursing mother in Montacute had sent a request of another "'to let her have her baby for awhile, 'cause Benny's mouth's so sore'" (*ANH*, p. 71). Finally, Clavers' own wardrobe is transferred to another body in an ominous doubling, the

neighbor Mrs. Nippers, having loitered in Clavers' home until invited to tea, though as she said, "she was hardly decent" (*ANH*, p. 144). Clavers proceeds graciously to lend her a "smart cap and cape, " a "pair of thin shoes, and then clean stockings...successively added as decided improvements to her array" (*ANH*, p. 144). The question to be raised in this mirroring is, of course, whether it is Mrs. Nippers who takes advantage of the opportunity to "level upwards," or whether Clavers' forbearance suggests a proclivity to "level down."

Private property, then, in *A New Home*, forms the boundaries of an inviolable space which is under constant assault in western society. And where Clavers suggests that communal property relations form a society of scarcity, since she declares that "one brass kettle is enough for a whole neighborhood," and can "point to a cradle which has rocked half the babies in Montacute," (*ANH*, p. 68) the reader is given to understand that the assault upon private goods is an assault upon the work ethic itself, labor being expended neither in the production of, nor to the end of purchasing, these goods. Moreover, while shared and expanded use-values (skillets, for example, perform double duty as wash basins) are linked directly to the reduction of labor necessary for individual acquisition, the ideology of *non-use* values as the markers of domestic economy and cultural refinement is non-existent on the frontier. Indeed, Clavers is plagued, in this "ruder stage[] of society, where no one has yet begun to expend anything for show," (*ANH*, p. 186) by her neighbors' consistent judgement of goods in terms of use, so that "a nest of japanned tables" is graded down to the status of firewood, Mrs. Jennings declaring "'they'll do better for kindlin's than anything else, here," (*ANH*, pp. 42-43) and a "tall-

cupboard" is turned out of doors to do "yeoman's service long afterward as a corn-crib" (*ANH*, p. 45). Within the regressive status of commodities, then, is inscribed the negative ethic of those who "despise plodding industry," and who "indulge in repining and feverish longings after unearned enjoyments" (*ANH*, p. 78).

If *A New Home* had, however loosely, structured for its readers a revised narrative of the virtuous woman assailed by moral (and material) indigence, it should come as no surprise then that Clavers should seek to construct within the narrow sphere of the domestic a sanctuary where one can keep those virtues intact. The facts of frontier life, however, had so well defined for her readers the imagery of a "fallen nature," that the mediating trope of sexuality germane to seduction novels could be dispensed with, and Clavers begins to structure an "incipient Eden" (*ANH*, p.79) as an economic fantasy whose example, she hoped, might transform Montacute society. For if a nagging question of the seduction novel had asked of the "rake" whether he might be reformed, Clavers had hoped for the best with regard to rakes male *and* female in Montacute. With this in mind, then, Clavers set about making of her home a model of cultivation and refinement based upon a domestic economy which could produce the material proof that Montacute, indeed, was "levelling up."

Such a project was, to be sure, a gendered one, since, "[a]s women feel sensibly the deficiencies of the 'salvage' (*sic*) state, so they are the first to attempt the refining process" (*ANH*, p. 147). However, where Sandra Zagarell has argued in her introduction to *A New Home* that gendered perspectives in the text suggest that masculine economic views "delineate[]... an economy devoted to profit making without material

productivity," while female economy is "based upon literal and life-sustaining production," [36] I would argue that Clavers' vision for a model economy is based upon an ideology of diversified consumption which attempts to carry Montacute society far beyond the confines of "life-sustaining production, " and away from an economics of scarcity delineated through an overriding attention to use-values. Framing the images of this economy within the sphere of domestic production and the home, she had sought, in a mode of representation typical of domestic fiction, to isolate this vision from the larger scale agrarian production whose project of bringing more and more land under cultivation had only meant for her a shifting and unstable set of social relations as families were uprooted.[37] The core image of this idealized economy is, she suggests, "the rough, pole-fenced acre, which we had begun to call our garden," (*ANH*, p. 113) and Clavers set to work, as she stated, "armed with the *American Gardener*, and quantities of choice seeds received... from home (*ANH*, p. 78) to "cultivate" this revision of "Michigan economics" (*ANH* , p.14).[38]

That Clavers should take up edenic imagery for a discussion of frontier gardening is hardly surprising. However, as a model of productivity, she had sought something more "deeply rooted" within the biblical narrative than a conventional icon. For Clavers had hoped to structure her economic vision squarely between the regressive images of an idle subsistence lifestyle and the excessive activity inscribed in the lapserian mandate to earn one's living by the sweat of the brow. Either end of this spectrum, after all, can connote the idea of scarcity, and she had perceived that both moral refinement and surplus production could

be figured within an ideology of well-earned leisure. Her aspirations to inculcate in her neighbors, for example, "taste enough to love what is so charming," she half-joked, might be "exemplified by one who sits *all day* in a carpeted parlor, " (*ANH*, p. 53) and Annette Kolodny reminds the reader that it was Clavers' economic situation which allowed her to enjoy the long walks and rides which are occasionally recounted in *A New Home.* [39] Indeed, it is after a lengthy digression upon the growing population of tradespeople and workers in Montacute that she remarks "[h]ere I am wandering like another Eve from my dearly beloved garden," giving the reader to understand that the discourse on labor constructs the margins, and not the core, of her sensibilities. (*ANH*, p. 80). One sees readily the contradictions implicit in this ideology, for it is only through the presence and labor of these classes that she could suggest the possibility of improvement and economic stability in the village to begin with. For the moment, however, the only activity constituting "a work of much time and labor" is the clearing, by means of "the immense 'breaking up' plough," the roots of shrubs and trees which Clavers declares the "troublesome proofs of the fertility of your soil" (*ANH*, p. 79). "But before this work is accomplished," she adds, "your incipient Eden will afford much of interest and comfort," and far form suggesting that domestic economy is defined through labor, gardening begins to constitute a "taste for [a] primitive source of pleasure"(*ANH*, p. 79).

There is implicit within the reference to "troublesome fertility," however, the equivocal image of nature unchecked, and the counter-claim to the merits of gentle cultivation as a force which can order and refine a chaotic landscape.[40] Clavers' images consistently link horticulture in the

west with a more diverse and ordered agriculture in the east, and she is quick to point out the prospect that "it is well worth while to make garden in Michigan, " (*ANH*, p. 82) for "[e]nrich it properly," she states, and "you need lack nothing that will grow north of Charleston" (*ANH*, p. 81). Moreover, Clavers carries this image of an ordered and rational productivity outside the bounds of the garden itself, suggesting that even the unbridled productivity of the wilderness can be controlled by the dictates of a domestic economy, uniquely figured in an image of barter with "the queen of the forest," whose offer to "'Schwap'" flour for gathered berries recontextualizes subsistence modes of production as paradigmatic of the work ethic and exchange (*ANH*, p. 81).[41] Both produce and producer, then, of the untamed wilderness, are drawn into Clavers' vision, and the equitable exchange of goods "in whatever vessel you choose, returning the same measured quantity" (*ANH*, p. 81) becomes a non-moneyed image of transfering surplus value (flour) into consumption for pleasure (berries). And though Clavers had had difficulty in getting her neighbors to "expend for show," she notes with some satisfaction that in dealing with Native Americans "[i]f you add to the price an old garment...or a string of glass beads, the treasure is at once put on and worn with such an air.... Broadway could hardly exceed it" (*ANH*, p. 81).

With the images of a fecund nature safely contained within the market relations of domestic economy, Clavers could then proceed to suggest that the constricted agricultural practices of her neighbors were in reality a fallen mode of production, defined, perhaps, in the rough images of clearings and the unvarying landscape produced by single crop farming. For she had situated, between her own practices and those of

her neighbors, the proof that nature itself revelled in "expending for show," and she had, moreover, been able to locate the forest Eve who attended this sphere. Visibility and diversity seem to be the key, then, to an edenic economy, and it is against this ideology that the reader is meant to gauge the fallen character of a class of people who, "in that terse and forceful style which is cultivated at Montacute," espouse the counter-ideology that "'Taters grow in the field, and 'taters is good enough for me'" (*ANH*, p. 82). Such a self-styled commitment to the"underworld" of agrarian activity inscribes this model of Michigan economy as a space neither domestic nor natural, and constitutes a zone of marginal production virtually indistinct from a retrograde lifestyle of root gathering. The levels of"cultivation" and character defined within differing modes of production, then, are readily measured against the fertility of Michigan soil, and one reads in Clavers' declaration that it "amply repays whatever trouble we may bestow upon it" (*ANH*, p. 81) the ethic which registers diverse consumption as the dividend of prudent labor.

It comes as little surprise, then, that Clavers should extend her vision of surplus productivity to the conspicuous foliage of flowers, and less that she should keep to the dictates of edenic economy in considering it "a sin" that "where the earth produces spontaneously such myriads of splendid specimens," one might forego "the little time and pains required for their cultivation" (*ANH*, p. 79). One notes again how readily the images of natural spontaneity and domestic economy are drawn together against the counter-claim of a neighbor who states "'she never know'd nobody make nothin' by raisin' sich things'" (*ANH*, p. 80). The moment is

significant because Clavers herself is hard pressed to find in response a locus of value in flora "exotic or home-bred" which can cut across economic contexts and "inspire [the] neighbors," as she states, "with some small portion of ... love for everything which can be called a flower" (*ANH*, p. 79). For if she had been committed to the idea of value within what was essentially a non-use mode of production, she could do so only in recognition of the fact that flowers somehow constituted the signification, rather than the presence, of real value. Such a transfer of real wealth into the "pure signs" of wealth is, as Jean Baudrillard has noted, marked through the capacity for "useless" expenditure, and Clavers implies as much in an abortive response to a woman to whom she offered a cutting of a "noble balm geranium" (*ANH*, p. 80).[42] "One might have enlightened her a little as to their moneyed value," she states, "but I held my peace and gave her some sage-seed" (*ANH*, p. 80). And though she understands well enough that such a view means little within this "ruder state" of Montacute economy, Clavers is quick to register her aspirations, and the incident is an occasion to furnish evidence of a burgeoning market for "show" on the frontier, for "oddly enough," she recalls, "there is, within three miles of us, a gardener and florist of no mean rank... whose aid can be obtained ... for some small consideration" (*ANH*, p. 80).

Clavers had realized, after all, that "small consideration" defined the scope of frontier women's willingness and capacity to generate disposable income, and yet it remained, to her sensibilities at least, the commitment to "some little extra personal exertion" which would enable "the introduction of those important nothings on which so much depends" (*ANH*, p. 147). "'Spinning-money,'" she suggested, "buys a looking-glass

perhaps, or 'butter-money' a nice cherry table" (*ANH*, p. 147). And while these acquisitions, she had hoped, along with the "[n]arrow beds... bright with Balsams and Sweet Williams, Four o'clocks, Poppies and Marigolds" (*ANH*, p. 147) might modestly transform the landscape of Montacute into a vision of refinement, she had hoped no less, that "some little extra exertion" on the part of her neighbors might alter the social landscape as well. For one notes in the images of women's productivity the market relations which held the goods Clavers herself depended upon, recontextualized, however, as a sphere of activity generated by frontier women themselves, made sensible to their own dependence on articles of domestic comfort. As a result, an ideology of needs equivocally registered through "important nothings" ultimately blurs the distinction between market and cultural refinements, and exposes in turn the intimate network of use and non-use values through which Clavers had indeed hoped to "cultivate" a stable vision of Montacute society.

After all, that Clavers should choose to articulate her desires within the reified images of commodity relations is hardly surprising, given the extent to which material concerns governed the daily life of frontier society as we see it in *A New Home.* Moreover, though I have suggested that Clavers' desires to refine her neighbors' homes can be linked to her own concerns to acquire goods and services monetarily, there is little question that her vision of a society stabilized through the acquisition of enduring goods is also a response to those economic practices which had most *destabilized* life in Montacute. For if Clavers had given her readers to understand that where families placed any emphasis at all in improvements to the home, "hers is the moving spirit," (*ANH*, p. 147) she

had also made it tacitly clear that this activity was framed by the more potent and literal "moving spirit" of males, whose attitudes toward economic improvement she envisioned as a phase of primitive accumulation, the drive for "the possession of a large number of acres" which "makes but little difference in the owner's mode of living" (*ANH*, p. 22). "[T]he habit of selling out so frequently," she stated, "makes that *home*-feeling... almost a nonentity in Michigan," since the man who "holds himself ready to accept the first advantageous offer , will not be very solicitous to provide those minor accommodations, which, though essential to domestic comfort, will not add to the moneyed value of his farm, which he merely considers an article of trade" (*ANH*, p. 22). So it is, then, that a rhetoric of small-scale acquisitions and "minor accommodations" had come to constitute, in *A New Home*, a model of "counter-productivity" against the large-scale and highly mobile practices of males on the frontier. For such a view at once acknowledges the limited sphere of excess productivity available to women, yet inscribes within its ethic of "extra personal exertion" the capacity to impress upon the home the markers of permanence and stability --- not only through the figures of commodities themselves, but through the practices and market relations which had been established to secure them.

"One must come quite away from the conveniences and refined indulgences of civilized life to know anything about them," (*ANH*, p. 183) Clavers would declare at the end of *A New Home*. Such sentiments are hardly novel in frontier writings, and yet the assertion remains striking in its radical shift of intention. For what had begun in earnest as "a sort of 'Emigrant's Guide'" (*ANH*, p. 1) for those heading west with the tide, ends

by turning its gaze once again to the east, and suggests that the true value of the text might lie in the fact that "all this forms part of the schooling which I propose for my spoiled child of refined civilization" (*ANH*, p. 185). It was, after all, "[t]his simplification of life, this bringing down the transactions of daily intercourse to the original principles of society," (*ANH*, p. 184) which had forced upon her the understanding that within the comforts of eastern life there existed "some absolute and evident superfluities, according to the... estimate of these regions" (*ANH*, p. 185). To "[the] doll of Fortune, who may cast a languid eye on this homely page," (*ANH*, p. 185) therefore, she had wished to impart some of the understanding that "the life-style that this ideal prescribed was accessible only to white, middle- and upper-class women in more settled areas of the country."[43]

It does little good to argue whether *A New Home* is , as David Leverenz has argued, "one of the most flagrant and insouciant instances of the patrician paradigm, surviving and at least temporarily prospering in the American wilderness," unless of course such argument is carried on outside the relative comforts of our own commitment to commodity culture and the complex web of economic relations which pervade modern life.[44] The Kirkland's, after all, did *not* wind up prospering in the Michigan wilderness, and the failed work of cultivating a society beyond its "original principles" must have seemed to Caroline Kirkland a feminine tragedy; the instability of frontier home-life was real enough, she understood, and cut equally across class-lines to form a core experience in women's lives.[45] And if it was the habit of middle-class women to gaze upon another whose labor they could hardly see in relation to their own lives as if "she may

have no shadow of claim... beyond that which arises from her being a daughter of Eve, (*ANH*, pp. 185-186) Mary Clavers asked that her readers "look at the matter from opposite points of view. *Her* light shows very plainly, as she thinks, what is *our* Christian duty; we must take care that ours does not exhibit too exclusively her envy and her impertinence" (*ANH*, p. 186). Caroline Kirkland suggested further that such a polarization might genuinely map the social geography of an expanding nation, and that in the "opposite point of view" that fashioned the eastward gaze of both "her envy and her impertinence," the woman who longed to make a true home in the wilderness sought only the "reflection of any of the features" of what had once been "her own dear fire-side" (*ANH*, p. 146).

---

[1] Caroline M. Kirkland, *Forest Life*. 2. vols. New York: C.S. Francis and Company, 1842. Reprint by Literature House: Upper Saddle River, N.J. , 1970. p. 34. Hereafter abbreviated as *FL*.

[2]Anonymous review, "*A New Home; Who'll Follow or, Glimpses of Western Life* By Mrs. Mary Clavers, an Actual Settler." *Knickerbocker Magazine* XIV (November, 1839), 452.

[3]Mary Kelley, *Private Woman, Public Stage: Literary Domesticity in Nineteenth Century America*. (New York: Oxford University Press, 1984), p. 129.

[4]Quoted by Edgar Allan Poe in *The Literati*. Vol. XV, *The Complete Works of Edgar Allen Poe*. ed. James A. Harrison. (New York: AMS Press, Inc., 1965. p. 85.

[5]Caroline M. Kirkland, *A New Home, Who'll Follow, or Glimpses of Western Life*. (New Brunswick: Rutgers University Press, 1990), p. 7. Hereafter cited as *ANH*.

[6]Irving had become separated from his party in *A Tour on the Prairies*, and responded in the following manner: "I now found myself in the midst of a lonely waste, in which the prospect was bounded by undulating swells of land, naked and uniform, where, from the deficiency of landmarks and distinct features, an inexperienced man may become bewildered, and lose his way as readily as in the wastes of the ocean.... To one unaccustomed to it, there is something inexpressibly lonely in the solitude of the prairie." Washington Irving, *A Tour on the Prairies*. (Norman: U of Oklahoma Press, 1956), p. 175. The distinction can clearly be seen as one which goes a long way toward articulating the divergent experiences of gender on the frontier, though it is not clear from Kirkland's own representations that all frontier women felt so constrained by the

lack of privacy. I would suggest, then, that the sentiment is class- structured as well.
[7]Quoted in William Osborne, *Caroline M. Kirkland.* (New York: Twayne Publishers, Inc., 1972), p. 44.

[8]Osborne's New York Lawyer went so far as to suggest that he residents of Pinckney had never actually read the book: "She keeps it out of Michigan....They told me Mrs. Kirkland intended to point out real characters in her neighborhood, and slander them most scandalously.... Not having seen the book, they are convinced from what they hear that they are all satirized and scandalized by name, almost." (Quoted in Osborne, *Caroline M. Kirklnd,* p. 45). While I doubt that Kirkland herself could have done much to keep the book out of Michigan, or that no one in Pinckney had read the text at all, the anecdote does nevertheless suggest an interesting perspective on readership and the circulation of written texts on the frontier.

[9]Henry Nash Smith, *Virgin Land: The American West as Symbol and Myth.* (Cambridge: Harvard UP, 1950), p. 219.

[10]Quoted in Nash Smith, *Virgin Land,* p. 219.

[11]See Osborne, *Caroline M. Kirkland,* for an historical overview of the family's "western experiment," esp. pp. 22-26.

[12]Kirkland chides that "[s]ome of my dear theorizing friends in the civilized world had dissuaded me most earnestly from bringing a maid with me" (*ANH* 39).

[13]Kirkland states in *Forest Life*: "We give our sincere, and hearty, and devoted support to that broad basis of our Constitution--- the natural equality of all--- but we are compelled to deny some of the inferences which are attempted to be drawn from it. The *social* equality of all is acted upon by nobody" (*FL*, p. 123). She sums up her argument by denouncing "the insincere and futile dogma that virtue, talent, accomplishments, manners, wealth--- should of right create no grades in social life" (*FL*, p. 124). Such privileges entail for Kirkland a call to civic duty, and the mandate to "counteract... the loud and hollow brawlings of the venal demagogue who pretends to long for agrarian laws, and to approve of universal and unmitigated coarseness" (*FL*, p. 124). Finally, in an apparent reference to de Toqueville, she complains that it "would seem to be left for foreigners to discern and explain the peculiarities of our position," (*FL*, p. 124) which in reference to the frontiers had meant to him that "[i]n this part of the American continent... the population has escaped the influence not only of great names and great wealth, but even of the natural aristocracy of knowledge and virtue. None is there able to wield that respectable power which men willingly grant to the remembrance of a life spent in doing good before their eyes." Alexis de Tocqueville, *Democracy in America.* Henry Reeve Text, ed. Phillips Bradley. (New York: Vintage Classics, 1990), p. 51.

[14]See J.G.A. Pocock, *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition.* (Princeton: Princeton UP, 1975), especially "The Americanization of Virtue," pp. 506-552. To bring Pocock's arguments more fully in line with Kirkland's own historical context, they should be particularly borne in mind with relation to the early debates on corporate charters and internal improvements, since 'self-interest' and the 'common good' are, as he suggests, the problem terms of an expanding commercial society in republican ideology. To some extent, the Kirklands must have felt this way about their own hopes for Pinckney, and the success of operations such as their grist mill.

[15]Joel H. Silbey, *The American Political Nation.* (Stanford: Stanford UP, 1991), p. 11.

[16]Michael Holt, *Political Parties and American Political Development from the Age of Jackson to the Age of Lincoln.* (Baton Rouge: Louisiana State UP, 1992), p. 26.

[17]For a general overview of Kirkland's attitudes toward Jacksonian Democracy, see John C. McCloskey's "Jacksonian Democracy in Mrs. Kirkland's *A New Home--- Who'll Follow*" *Michigan History* XLV (December, 1961), 347-52.

[18]For an interesting discussion of contested partisan claims to republican values in the early Nineteenth- Century see Charles Sellers, *The Market Revolution: Jacksonian America 1815-1846*. (New York: Oxford UP, 1991), esp. "Ambiguous Republicanism," pp. 34-69, and "Ambiguous Democracy," pp. 332-364. It would certainly be problematic to suggest here that since Kirkland's political views appear to run counter to those of the Democrats, she was aligned with the Whig party instead; Clavers, at any rate, is aware that her political statements are qualified by what she calls "my feminine capacity merely," (*ANH*, p. 48) as a woman without suffrage. However, it seems likely she would have understood that in remarks given by Webster in New York in 1831 lay a vision of commerce and internal improvement which would have done much to alleviate her difficulties. Speaking of an expanding frontier, he appealed to his audience "to remember that new States have arisen... spreading and thickening over vast regions which were a wilderness when the Constitution was adopted. Those States are not... directly linked with maritime commerce. They are entirely agricultural, and need markets for consumption; and they need, too, access to those markets. It is the duty of the government to bring the interests of these new States into the Union, and incorporate them closely in the family compact." *The Works of Daniel Webster*, Volume I (Boston: Little, Brown and Co., 1858) , "Public Dinner at New York," pp. 207-208. Clavers' constant depictions of the poor state of roads "but little worked,... and in some parts... almost in a state of nature" (*ANH*, p. 34) may be read as political statements in themselves; however, it will be seen that where partisan platforms such as Whig banking policies work to disrupt the domestic spaces of frontier women, political alignment quickly gives way to gender politics of male vs. female activities.

[19] Nancy Grey Osterud, "Gender and the Transition to Capitalism in Rural America," *Agricultural History* 67(2) 1993, pp. 14-29, cited p. 17.

[20] The depiction of a diverse economic network evoked in this passage is useful in its representation of exchange, though it should not be conflated with an over-optimistic index of market involvement among frontier women. Rather, one finds that such activity is best situated between a perhaps idealized vision of radical self-sufficiency, and that of a pervasive "market consciousness." Winifred Barr Rothenberg distinguishes between the economic system of the "market-place," and that of a "market economy," noting that the former, as developed, and localized networks of various forms of exchange have existed since earliest civilization. On the other hand, it is "the articulation *between* market-places that defines an integrated market economy," and which tends to shift individuals' practices more and more fully to production solely for profit which will buy goods in a developed commodity market. The latter tends to be felt as an external series of forces governing the costs of goods and services over broad sectors of the economy. See Winifred Barr Rothenberg, *From Market-Places to a Market Economy: The Transformation of Rural Massachusetts, 1750-1850*. (Chicago: The University of Chicago Press, 1992) quoted p. 20. The issue is of crucial interest here, since a degree of market orientation broadens the scope of women's productivity beyond the confines of a model of subsistence; for although it has been typical to view non-market networks of exchange among women as governed by a more humane and communal set of values (what Rothenberg calls the "moral economy"), Alan Kulikoff has argued effectively the patriarchal maintenance of control within subsistence agriculture, thereby governing the consumption of women's productivity, and Nancy Grey Osterud has suggested that while the commercialization of crops tended to marginalize women's contributions to family income, what had been come to be seen as "petty production," such as the sale of butter and eggs evolved for farm wives into a sphere of "income from these operations virtually invisible to their husbands." See Alan Kulikoff, *The Agrarian Origins of American Capitalism*. (Charlottesville: University Press of Virginia, 1992), esp. "The Transition to

Capitalism in Rural America," pp. 13-33, and Grey Osterud, "Gender and the Transition to Capitalism in Rural America," quoted p. 22. One may, then, argue the 'morality' of market-behavior in women while acknowledging the communal values implicit in this depiction of exchange.

[21] Clavers recounts the spurious gossip of a Mrs. Nippers at a meeting of the Montacute Female beneficent society, for example, who states dissatisfaction "that those who lived in *log-houses* at a little distance... had not been invited to join the society" (*ANH*, p. 136), though another neighbor charges that it was Mrs. Nippers herself who had gone "round her neighborhood last Friday, and told how that Mrs. Clavers and Miss Skinner despised everybody that lived in log-houses" (*ANH*, p. 136). See Annette Koldny, *The Land Before Her: Fantasy and Experience of the American Frontiers, 1630-1860.* (Chapel Hill: University of North Carolina Press, 1984), esp. "The Literary L legacy of Caroline Kirkland," pp. 131-158 for a good discussion which often links domestic structures in *A New Home* to the images of class ideology and levels of moral/cultural refinement.

[22] Raymond Wiliams, *The Country and the City.* (Oxford: Oxford UP, 1973), p. 26.

[23] The passage forms an interesting breach in what V.N. Volosinov has argued is a containing, and controlling strategy at the core of the framed, or "reported speech" of others. He states in this context that "[l]anguage devises means for infiltrating reported speech with authorial retort and commentary in deft and subtle ways. The reporting context strives to break down the self-contained compactness of the reported speech, to resolve it, to obliterate its boundaries. We may call this style of speech reporting *pictorial*." V.N. Volosinov, *Marxism and the Philosophy of Language.* tr. Ladislav Matejka and I.R. Titunk (New York: Seminar Press, 1973), p. 120. The last statement is certainly germane with respect to Clavers' understanding of the politics of "picturesque behavior."

[24] See Brigette Bailey, "Irving's Italian Landscapes: Skepticism and the Picturesque Aesthetic," in *ESQ* 1986 (Spring) 32 (1): 1-22, quoted p. 2. Also, particularly relevant with respect to the sketch in the expanding west, see Kris Lackey, "Eighteenth-Century Aesthetic Theory and the Nineteenth-Century Traveller in Trans-Allegheny America: F. Trollope, Dickens, Irving and Parkman," in *American Studies* 1991 (Spring) 32 (1): 33-48.
[25] See Kulikoff, "Free Migration and Cultural Diffusion in Early America, 1600-1860," in *The Agrarian Origins of American Capitalism*, pp. 183-225, for a good discussion of patterns of migration in this period which were far more complex than can be explained by a model of western opportunism. For a discussion of hard times in urban labor markets during this period, see Charles Sellers, *The Market Revolution*, esp. "The Burgeoise Republic," pp. 364-395.

[26] There are moments in William Osborne's reading of *A New Home* which suggest the pervasiveness of this "incomprehensible" view of scarcity in the land of opportunity. Commenting upon the Kirkland's impatience with the widespread practice of borrowing on the frontier, he states that "[t]o the Kirklands, who had not set out for the West until they had assured themselves that their needs were provided for, such a habit was outrageous and personally annoying." Osborne, *Caroline M. Kirkland*, p. 46. The fact that the Kirklands looked to the west as entrepreneurs with capital *and* land at their disposal needs to be examined more closely.

[27] Sellers, *The Market Revolution*, p. 27.

[28] Winnifred Barr Rothenberg notes in Massachusetts farm books "the appearance of labor contracts by the mid-eighteenth century and their increased frequency after 1800." Her examination of the implied obligations between parties negotiating contract labor suggests that the security offered to both employer and employee was viewed as mutually beneficiary, though either side may have sought out contract labor as primarily

self-protective behavior. See *From Market-Places to a Market Economy*, esp. "Contract Labor in Massachusetts Agriculture," pp. 181-212, quoted p. 183.

[29] Sandra L. Myres notes that "abortion was readily available during the middle-decades of the nineteenth century, and several recent studies have suggested that upper and middle-class married women as well as unmarried and poor women used abortifacients... or engaged the services of abortion practitioners." Sandra L. Myres, *Westering Women and the Frontier Experience: 1800-1915.* (Albuquerque: University of New Mexico Press, 1982), p. 155.

[30] Charles Sellers argues that libidinal repression in the early nineteenth century amounted to "[c]apital conscripting human labor to productive manipulation of inert matter," and that for "a bourgeois leveling the wilderness, fighting off alcohol, fleeing dirt, and draping its carnivorous animality in ceremonious table manners, sexuality epitomized the uncontrolled nature they had to subdue." Sellers, *The Market Revolution*, p. 259. The pertinence of these images in relation to *A New Home* would be hard to gloss.

[31] Hector St. John de Crèvecoeur, *Letters From an American Farmer.* (New York: E.P. Dutton & Co. Inc., 1945), pp. 46-47.

[32] *ibid.*, p. 47.

[33] For a specific discussion of land speculation in Michigan as relates to *A New Home,* see John C. McCloskey, "Land Speculation in Michigan in 1835-36 as Described in Mrs. Kirkland's *A New Home-- Who'll Follow?*" *Michigan History* XLII (March, 1958), pp. 26-34. For a more general overview of United States land policy during this period, see Paul W. Gates, *Landlords and Tenants on the Prairie Frontier: Studies in American Land Policy.* (Ithaca: Cornell UP, 1973), esp. "The Role of The Land Speculator in Western Development," pp. 48-71.

[34] See "The Morality of Improvement" in Williams, *The Country and the City*, pp. 60-67. It should be noted here that Williams, in *The Country and the City* engages solely the history of agrarian production in Great Britain, and that there are, to be sure, major differences in American agricultural history that suggest caution in simply "applying" his arguments to conditions on this side of the Atlantic. However, it is as much his intention to chart the discursive practices which have reflected and refracted this history as it is to graph its development, and the text has, in my view, a wealth of information with regard to agriculture and literary representation which is pertinent to *A New Home.*

[35] Cathy Davidson, speaking of one such novel of seduction, *The Power of Sympathy,* states that "seduction... is a metaphor not just of women's status in the Republic, but of a range of problems, all of which might be reduced to the same structure of the seduction plot--- that is, a range of problems that arise when moral value and social responsibility are outweighed by the particular desires... of privileged individuals or classes." The ideology of privilege may require inversion, as we have seen, on the frontier, but the moral implications remain, in my opinion, the same. *Revolution and the Word: The Rise of the Novel in America.* (New York: Oxford UP, 1986), p. 108.

[36] Sandra Zagarell, "Introduction," to *A New Home, Who'll Follow,* pp. xl-xli. While I tend to agre with her overall discussion of *A New Home,* Zagarell's arguments are at times problematized by the assumption that land-speculation is the only sphere of profit economy in the west, and that subsistence economics dominates all other spheres of production delineated in the text. This is an over-simplification of the history of American agricultural society, and Joyce Appleby, for example, has suggested that even in the early Republic the market may indeed have been a benign force in American farmer's lives, arguing that "population pressure in Europe had raised the price of food

and the incentive of higher grain prices had encouraged American farmers to plan for ever-larger surpluses." Joyce Appleby, *Capitalism and the New Social Order: The Republican Vision of the 1790s*. (New York: New York University Press, 1984), p. 98. Clavers' own image of the distillation of wheat into whiskey as a "branch of business" (*ANH* 7) certainly belies Zagarell's statement that wheat forms a central core of "literal and life-sustaining production." p. xl. Moreover, while *A New Home* is certainly sparse in its representations of larger-scale agrarian production, this has as much to do, in my opinion, with the limited sphere in which Clavers suggests women move, as with the implication that nothing of the sort was going on. When, for instance, Clavers recounts having woken up at the Danforth's unsure of the time of day, she is told by "one of the 'gals'....'why the boys has been gone away this hour'" (*ANH*, p. 9). These "lords of creation," as Clavers calls them, return to eat their dinner, and she states that "as each one finished , he arose and walked off" (*ANH*, p. 14). It is, therefore, the wasteful expenditure of surplus production, and the complex factors limiting access to commercial domestic goods which best defines Clavers' problematic stance in relation to market behavior, and hardly its simple *presence*.

[37] See ,for example, Gillian Brown's "Getting in the Kitchen with Dinah: Domestic Politics in *Uncle Tom's Cabin*, in *Domestic Individualism: Imagining Self in 19th Century America*. (Berkley: University of California Press, 1990).

[38] The image of the garden in women's frontier writing forms the organizing principle of Annette Kolodny's *The Land Before Her*, and is the most comprehensive examination of garden imagery in relation to this field of writing. While I tend to disagree with her assertion that the garden is for women "some part of the landscape otherwise physically appropriated by men for the marketplace," (for I wish to argue that the garden can be viewed as an idealized "feminine market" based on productivity and diversified consumption) the garden does, in my opinion, form the space where women's "imaginative play... focused on the spaces that were truly and unequivocally theirs: the home and the small garden." *The Land Before Her*, quoted respectively, p. 12, and pp. 5-6.

[39] See Kolodny, *The Land Before Her*, pp. 145-46.

[40] See especially Milton's *Paradise Lost*, Book IX, where Eve remarks of the garden that "what we by day/ Lop overgrown, or prune, or prop, or bind,/ One night or two with wanton growth derides/ Tending to wild." Book Ix: 209-12.

[41] Marx's following statement in the *Grundrisse* is of interest here: "Exchange between different tribes or peoples-- and this, not private exchange, is its first form-- begins when an uncivilized tribe sells (or is cheated out of) an excess product which is not the product of its labor, but the natural product of the ground and of the area which it occupies." Karl Marx, *Grundrisse: Foundations of the Critique of Political Economy*, tr. Martin Nicolaus (New York: Penguin Books, 1973), p. 170.

[42] See especially "The Art Auction," in Jean Baudrillard, *For a Critique of the Political Economy of the Sign*. (Telos Press, 1981). pp. 112-122.

[43] Barbara J. Harris, *Beyond Her Sphere: Women and the Professions in American History*. (Westport, Connecticut: Greenwood Press, 1978), p. 33.

[44] David Leverenz, *Manhood and the American Renaissance*. (Ithaca: Cornell UP, 1989), p. 152.

[45] Annette Kolodny notes that "[d]issafected, disappointed, and having lost much of their initial investment, the Kirklands left Michigan for New York City in 1843." Kolodny, *The Land Before Her*, p. 154.

CHAPTER 3

"A REMARKABLE LOAD OF ORIGINAL SIN": REDEEMING THE PROFANE LANDSCAPE IN YELLOWSTONE NATIONAL PARK[1]

In a journal entry for Monday, August 29, Nathaniel Langford, member of the 1870 Washburn-Doane expedition to the headwaters of the Yellowstone, paused to reflect on the "taste evinced by our company in the selection of names for the various objects we have met with." Langford, writing about the gorge on Cascade Creek given the name of the Devil's Den, referred to the party's habit of naming Yellowstone features in accordance with what their senses (and some cultural baggage) suggested to them as they surveyed the unearthly landscape of the region. He justified the practice by stating that the objects were:

> all so different from any of Nature's works that we have ever seen or heard of... and withal so full of exhibitions which can suggest no other fancy than that which our good grandmothers have painted on our boyish imaginations as a destined future abode, that we are likely, almost involuntarily, to pursue the system with which we have commenced, to the end of our journey.[2]

Yet it was perhaps the peculiar Mud Volcano, that most objectively named of Yellowstone feature's, which most challenged Langford's faith in a comprehensible landscape, and, if we are to take him at his word, in the credulity of his own experiences. Located within what was to be the nation's first national park, and destined for bad press by a host of future visitors, one of whom described it "the most disgusting object in nature," this feature so struck Langford that he felt compelled to visit it twice.[3] His entry for September 3 states that he returned "to see it again for the one especial purpose... of assuring myself that the notes made in my diary a few days ago were not exaggerated." And though Langford could boast

that "No! they are not!," his entry offers little assurance that he truly felt the master of his senses, or of the scene which lay before him. The volcano itself remained problematic, for he noted further that:

> The sensations inspired in me today, on again witnessing its convulsions, and the dense clouds of vapor expelled in rapid succession from its crater, amid the jarring of the earth, and the ominous intonations from beneath, were those of mingled dread and wonder. At war with all former experience it was so novel, so unnaturally natural, that I feel now while writing and thinking of it, as if my own senses might have deceived me with a mere figment of the imagination. But it is not so.[4]

Wavering between the ground of raw nature, and the groundwork of experience, the "unnaturally natural" landscape of Yellowstone seemed equivocally sublime, at best; at its worst, it seemed "at war" with the very habits of observation and understanding.[5] Facing such a scene, Langford would be one of the first, yet hardly the last, to argue he had doubts about putting pen to paper in an act of honest description.

Yet Langford's rhetoric of self-doubt is perhaps best understood as a cautious appeal to an American public whose knowledge of the Yellowstone region had been colored by such legendary figures as Jim Bridger, who along with other mountain men such as John Colter and Osborne Russell, remained one of a handful of white explorers who claimed to have visited the area. Bridger's propensity for the tall tale preceded him, and park historian Hiram Chittendon noted that as "the greatest romancer of the West in his time," his "reckless exaggerations won for him a reputation which he could not shake off when he wanted to."[6] His stories of cooking fish on the line by drawing them slowly through successively hotter strata of Yellowstone Lake, of forests in which "[a]ll forms of life" could be seen "standing about in stone where they were suddenly caught by ... petrifying influences," and of firing ineffectually at

the mirror image of a distant elk in a "mountain of perfectly transparent glass," seemed to wrap the region in the aura of a landscape inherently deceptive, and contrary to all sensible experience.[7] While there were few who would question Bridger's extensive knowledge of the west, or his skills as a guide, he seems to have been frustrated in his attempts at convincing the public that such a place as Yellowstone really existed, since "no periodical or newspaper would lend itself to his service."[8]

Langford, however, had more than his own reputation to be concerned about when it came to disseminating information as to what the party had seen in Yellowstone. Aubrey Haines notes that Langford was at the time "a nominal employee of Jay Cooke & Company, agents for the Northern Pacific Railway," and that he "probably promoted that exploration as a means of fulfilling his commitment to provide publicity."[9] The Northern Pacific, which was to play a consistent role in the development of the park as a tourist attraction, was eager to build a branch line to the area as a means of attracting passengers for its main east-west route.[10] And although as recently as the previous year explorer David Folsom had told Langford that, with regard to relating the events of the 1869 Cook-Folsom trip into Yellowstone, "he was unwilling to risk his reputation for veracity, by a full recital, in the presence of strangers, of the wonders he had seen," members of the 1870 trip published and spoke with what must have appeared a wanton disregard for their own good names.[11] Cornelius Hedges, Walter Trumbull, Warren Gillette, and Henry Washburn all wrote articles for papers such as the *Helena Herald* and Denver's *Rocky Mountain News*, with Washburn's account eventually finding its way to the *New York Times*. Langford himself wrote an account for *Scribner's Monthly* before publishing his journal.

The best measure of the party's success in creating a credible account, however, lies with the events precipitated by Langford's first lecture in a series for Cooke's publicity campaign, presented January 19, 1871 at the Lincoln Hall in Washington, D.C.. Present in that audience was the head of the U.S. Geological Survey of the Territories Ferdinand V. Hayden, who, inspired by Langford's lecture would make the first government sponsored expedition to the Yellowstone region later that same year.[12] It would appear, then, that Langford and the others managed to avoid the frustration Bridger was said to have felt at the hands of an incredulous public; at any rate, the Hayden expedition would in due time bring back a mass of scientific data concerning the park's features, and more important, perhaps, a wealth of sketches and photos done by artist Thomas Moran and photographer William Henry Jackson. If indeed the Hayden expedition of 1871 had an "aim ... to discern fact from fiction, once and for all," the American public consumed the images of Yellowstone with a zeal that betrayed a burgeoning faith in the wonders of the far western landscape.[13]

Chittendon, at least, seems to have understood that the novelty and spectacle of the west evoked experiences for which Bridger's brand of hyperbole seemed apt. Discussing the legendary tales of this guide and mountain man, he suggested that so long as such accounts "all have a basis in fact," they harbor "the 'soul of truth' which a great philosopher has said 'exists in things erroneous'.... [I]t is a fortunate thing," he argued further, "that these stories grow and develop with time, gravitating always from the real to the ideal; and he is to be pitied who feels an unseemly anxiety for the basic facts or would rob them of a single increment which the rolling years have given them."[14] To be sure, the

great landscape painters on the American scene in the nineteenth century expressed a similar penchant for capturing the ideal in western scenery. In the works of artists such as Frederic Church, Albert Bierstadt, and Moran as well, Barbara Novak discerns a tension between realistic detail and an idealized overall effect which she argues lay at the heart of a Christian model of the sublime, creating a landscape that was charged with the task of unfolding nature as "a holy text which revealed truth and also offered it for interpretation."[15]

And yet for those who felt themselves limited to the strictures of language in their accounts of the west and of Yellowstone in particular, and who found themselves as had Langford confronting scenes "at war with all former experience," the fear of failure, rather than a faith in revelation, would shape the conventions of written description. Challenged in their attempts to convey impressions of a landscape which appeared to be creating and destroying itself in successive moments, whose thermal springs led Hayden to abandon his geologic lexicon to remark on "the exquisite beauty of the coloring and the variety of forms" which " baffles any attempt to portray them, either with pen or pencil," a landscape whose sights blended radically with odors, sounds, heat, and texture, early visitors to Yellowstone would consistently share with Margaret Cruikshank the resolution that "[i]t is indeed wonderland and mortal language seems entirely inadequate."[16] In the years following the 1872 Act of Congress for the creation of Yellowstone, the National Park was to play host to American presidents, well-known authors, conservationists, and visitors, all who wrote from a variety of impulses; however they chose to do so, however poorly or well they managed, the Yellowstone landscape touched upon deep cultural assumptions they held

about nature and natural processes which inhered in the fabric of language itself.

That the park should strain the conventions of descriptive language for the nineteenth century traveler did not necessarily set it apart from other regions of the far West whose "singular qualities," Anne Farrar Hyde argues, "horrified most observers because... their education and experience had given them no words or images to describe what they saw."[17] Yet it was Yellowstone's unearthly, often disarming show of natural force which made it unique in the Western landscape, and if early explorers and promoters had become adept at glossing less appealing aspects of American scenery in their descriptions of other regions, the fact was that the geysers and springs of Yellowstone formed the central features of the area, and as a result shaped the primary aesthetic responses of observers. At one level, the sensual complexity of the landscape itself seemed overwhelming; the observer had little choice but to experience the park as a tumult of sights, smells, sounds, and thermal impressions which were in themselves often frightening. Carrie Adell Strahorn managed her description of a spring near Norris Geyser Basin by stating that:

> The hole looked like the very entrance to Hades, and the groans and hisses seemed the diabolical laugh of Pluto and his imps giving the mythical degrees of torture to his victims. The blackness of night was upon every rock, and through the dismal, darkened dive came only the murky stream laden with the sickening fumes of sulphur from an overheated cauldron.[18]

At a far more complex level, however, and one to which the above description is related by its profane imagery, lay the problem of the role such a landscape was to have in fashioning a suitable iconography for a culture which gave immense currency to its natural features. If, indeed,

the North American landscape could be read as a holy text in the Christian tradition, then American citizens would have it that in the possession, transformation, and ultimate redemption of the landscape, lay a revelation of providential history at work shaping the culture itself. Barbara Novak speaks of the "self-image" inherent in "these projections on nature" which "augmented the American's sense of his own unique nature, his unique opportunity," and which "could indeed foster a sense of destiny" characteristic of the nationalism shared by nineteenth-century Americans.[19] Readers of national literature consumed images of the natural landscape which had become heavily invested with the teleological rhetoric of the Christian Republic; characteristic are James Brooks' claims in an 1835 essay in the Knickerbocker that:

> God has promised us a renowned existence, if we will but deserve it. He speaks this promise in the sublimity of Nature. It resounds all along the crags of the Alleghenies. It is uttered in the thunder of Niagara. It is heard in the roar of two oceans, from the great Pacific to the rocky ramparts of the Bay of Fundy. His finger has written it in the broad expanse of our Inland Seas, and traced it out by the mighty Father of Waters! The august TEMPLE in which we dwell was built for lofty purposes. Oh! that we may consecrate it to LIBERTY and CONCORD, and be found fit worshippers within its holy wall![20]

Christian destiny and more secular visions of the Republic would not be so easily united throughout the whole of the nineteenth century, yet nature's figurative richness meant that it would mediate and maintain ties with both.

The public fascination with Yellowstone, however, could hardly have rested upon such prophetic images of God's and America's Nature. To be sure, the obvious volcanic activity of the area depicted for many visitors such as Charles Dudley Warner "a process of creation... going on daily and yearly," a process which suggested the grand design of Nature in the "slow formation of terraces and mounds and well-curbs, all

exquisitely sculptured, now like lace-work, now like the chiseling of a sculptor." But Warner expressed a concern as well, that the very forces at work in creation could appear as thoughtless as they had intentional, since the observer with a sense of the geologic past feared that "these lovely things" had been "created only to be destroyed in a great upheaval of the internal forces."[21] Geothermal features in Yellowstone appeared to lack the characteristics which could inspire visions of national progress, religious *or* secular. Warner was only one of a host of visitors who would register an anxiety regarding the Mud Volcano's endless and "fascinatingly loathsome" fate in that "this awful mixture cannot get out, and the creature has to swallow it again, and is perpetually sick to nausea."[22] In this respect, at least, Yellowstone was truly "at war" with the nation's ideas about nature. The region seemed best expressed in the terminology of the profane, rather than the sacred, and it is characteristic that from within the prevailing values surrounding American landscape, Yellowstone nomenclature and imagery so often found its way to the unregenerate world of the Fall.

Yet almost immediately after the expeditions at the opening of the decade, Yellowstone was made a national park in 1872, "dedicated and set apart as a public park or pleasuring-ground for the benefit and enjoyment of the people."[23] An apocryphal account circulated widely of a conversation held by members of the Washburn-Doane expedition at the junction of the Firehole and Gibbon rivers; the story had it that the party resolved there to sacrifice the opportunity to capitalize on their discoveries, in favor of bequeathing to the citizens of a democratic nation the spectacle of Wonderland.[24] Those citizens would eventually arrive, as well as visitors from around the globe, in comfortable railway cars, only to be whisked

into not so comfortable stagecoaches for the standard tour of the sights. Collectively, they would evoke the imagery of Dante, of Revelations, and of Pluto's underworld, all of which seemed a pity to one driver who remarked to Rudyard Kipling that it was a "cur'ous kinder folk" who would "come so far just to liken Norris Basin to Hell."[25] And finally, many of these tourists would find themselves at moments dismayed in a landscape "covered with a grey-white sepulchral deposit that is ghastly," where "clouds of vapor hang shroudlike above it," and "the earth trembles with a strange rumbling... and all vegetable life is extinct," a landscape which, in short, "emphasizes the ghastly look of death and destruction."[26] If Yellowstone were truly a national playground, it remained an enigma in the national landscape.

This is far too sweeping an indictment of Yellowstone's scenery, however. The Grand Canyon and Falls of the Yellowstone river, for example, contained many of the standard elements of the sublime, and visitors overlooking the great chasm were properly "appalled and fascinated, afraid and yet compelled to cling there" as they "mark[ed] the sculpturing of God upon those vast and solemn walls." Its geological features could be appropriately considered the "prototype" of European architecture, and then "of course, and almost beyond all else," one was "fascinated by the magnificence and utter opulence of color.... The whole gorge flames," wrote Dr. Wayland Hoyt in this often repeated and reprinted description, "[i]t is as if the most glorious sunset you ever saw had been caught and held upon that resplendent, awful gorge."[27] Yellowstone Lake offered as well a more subdued and picturesque landscape to visitors such as Mary Bradshaw Richards, who noted in her diary a reluctance to leave "this enchanting shore. The pure, exhilarating

atmosphere," she stated, "its savage solitude, its scenery of varied, inexhaustible beauty, have given us a season of content and rest kin to the 'Peace that passeth understanding."[28]

Yet this duality inherent in Yellowstone's landscape made the region seem less a model pleasuring ground than a contested space in a symbolic arena which had become familiar to Americans in their attitudes toward nature. Speaking in terms of aesthetic values, Yellowstone's landscape appeared to elicit what Roderick Nash has called a "double-mindedness" in the observer who could be both drawn to, and repulsed by, the experience of wilderness. Untouched nature was at once beatific in its revelations of God's creative forces, yet lacking in its absence of human artifice, and frightening in its potential for disorder.[29] This "double-mindedness" mediated the values of conquest and reverence with respect to nature, a mediation which ultimately shaped a vision of the ideal, civilized landscape as it unfolded on the frontiers. Moreover, as these values acquired force in shaping a complex vision of historical processes at work in the nation at large, and the frontier in particular, landscape itself accrued value insofar as it held the potential to transcend the moment in a sort of *nunc stans*, allowing the observer to judge both the past and future of human affairs in one expansive vista. Seeing both transacted and untransacted destiny in the sights of Montana Territory, for example, the Reverend Edwin J. Stanley paused before crossing into the Yellowstone Valley, and looked back upon the valley of the Gallatin River "flecked with herds of cattle, farm-houses, ranch cottages, and vast fields waving with golden grain ripe unto the harvest." He was, as it were, poised between two very disparate moments in the drama of national expansion, and yet the backward glance held the force of futurity, evoking

in Stanley a characteristic mingling of Biblical and Republican optimism. He freely stated his pleasure that:

> Enterprise and thrift had made this wilderness to bloom and the desert to rejoice and blossom as the rose. Advancing civilization has already planted here and there the nuclei of 'future great' cities, and relieves the dull and monotonous outline of swelling foot-hills, making bright contrasts in the distant landscape.[30]

One notes clearly the aesthetic impulse giving way to the forces of civilization as the landscape acquires value primarily in terms of its ability to map the topography of progress guided by human effort and divine prophecy.

It seemed as if Yellowstone could have escaped such a fate, however, for clearly the region held out little hope for the redemptive project such as it had been articulated in the ideology of manifest destiny, and the reclamation of the "desert" west. Indeed, John Sears has noted that the promotion of the park bill to Congress was prefaced by statements in the 1871 report of the House Committee on the Public Lands to the effect that "as far as agriculture and mining were concerned, the Yellowstone region was a wasteland."[31] In essence, Yellowstone's future as a "pleasuring-grounds" was assured in part because it was so unfit for the redemptive arena of human work; even so staunch a naturalist as John Muir would echo the House Committee report at the end of the century, suggesting the prevalence of use values with regard to public lands. Muir assured his Atlantic Monthly audience that:

> The withdrawl of this large tract from the public domain did no harm to any one; for its height, six thousand to over thirteen thousand feet above the sea, its thick mantle of volcanic rocks, prevent its ever being available for agriculture or mining, while on the other hand its geographic position, reviving climate, and wonderful scenery combine to make it a grand health, pleasure, and study resort,--- a gathering-place for travelers from all the world.[32]

If, then, Yellowstone had no working future in the designs of expansion, at least it held out the opportunity to showcase American character and scenery. Moreover, if the region could not be harvested for natural resources, it was clearly fertile territory for the visionaries of a growing leisure culture who, like Stanley, looked to "future years, when railroads shall checker this mountain country, and myriads of excited men and women shall come from every clime in splendid palace-cars to behold the curiosities of this wonderland."[33]

The Reverend was betraying a sense of uneasiness in regard to the park's future, since he was glad to have "made the grand rounds on a cayuse pony," a mode of transport which, if fitting to adventurous spirits, was also trying.[34] But in general, Americans took to their leisure in Yellowstone with a desire for whatever amenities could be offered throughout its early development. In growing numbers, they sought relief from the pressures of an urban-industrial society as part of a Victorian "duty to play" which, as Glenn Uminowicz argues, was "added to the stolid virtues of the Protestant work ethic--- industry, sobriety, and thrift," in part to stem the concern of clergy that Americans not lose track of religion in the midst of worldly success and the drive to achieve it.[35]

Indeed, it was Nature, not the tourist, that visibly labored in the park. John Muir saw in Yellowstone's active forces an image of:

> Nature working with enthusiasm like a man, blowing her volcanic forges like a blacksmith blowing his smithy fires, shoving glaciers over the landscapes like a carpenter shoving his planes, clearing, ploughing, harrowing, irrigating, planting, and sowing broadcast like a farmer and gardener, doing rough work and fine work, planting sequoias and pines, rose-bushes and daisies; working in gems, filling every crack and hollow with them; distilling fine essences; painting plants and shells, clouds, mountains, all the earth and heavens, like an artist, --- ever working toward beauty higher and higher.[36]

For Muir, too, these processes contained the force of Revelations--- "a look into a few of the tertiary volumes of the grand geological library of the Park" allowed the tourist to "see how God writes history."[37] And while it may well be that Muir's evocation of nature as a feminine agency performing God's handiwork implies complexities that he himself had not thoroughly considered, there is little doubt that the work of divine history progresses within a sphere of male prerogative. In her best moments, Nature works "like a man;" the teleology inherent in this activity takes its force from those primarily masculine images of labor which had, in the nineteenth-century mind, so clearly lent purpose and permanence to the work of westward expansion. The culmination of all this enterprise is an aesthetic impulse which is hard to overlook in its progressive ideology. "Rough work" leads to "fine work," in other words, and the creative/civilizing project evolves into a set of creative acts whose end is the sheer contemplation of beauty.

Yet the "liveliness and intensity of Muir's response," argues Sears, "appears... in part to be an attempt to compensate for the disturbing, destructive, and infernal aspects of Yellowstone's landscape."[38] Even in his most optimistic moments, when from a distance Yellowstone's thermal features spoke not as "lawless cataclysms," but as "shocks and outbursts" which "each and all tell the orderly love-beats of Nature's heart," Muir retained the sense that this was still a "transient, shifting show of things called 'substantial,'" and that the mortal visitor to the park had a fleeting experience at best in what might be "heaven and the dwelling-place of the angels."[39] Indeed, for many visitors to Yellowstone this sense of the ephemeral bespoke what was most disconcerting in the forces of nature. Rudyard Kipling, projecting his own uneasiness upon the common

understanding that Native Americans held a superstitious dread of Yellowstone, responded oddly to the colorful Fountain Paint Pots by stating "[d]o you wonder that in the old days the Indians were careful to avoid the Yellowstone? Geysers are permissible, but mud is terrifying." The Paint Pots are hardly one of the more "potent" features in Yellowstone; yet this reduction in the scope of natural forces, from the explosive to the vulgar and unassuming, nonetheless seemed to mark for Kipling a salient insight into the transience of experience itself, for he noted his amazement that an "old lady from Chicago took a piece of it, and in half an hour it died into limedust and blew away between her fingers. All maya--- illusion--- you see!"[40]

Moreover, the forces at work in Yellowstone which had so clearly made, and remade the landscape, could create a disarming sense of the relationship between humans, and the natural environment in which they sought to create a suitable, and permanent place of habitation. Kipling's description of futility even extends itself comically to "a road across a level park hotly contested by the beavers," dragging government maintenance into a seasonal cycle of fleeting works, since "every winter [the beavers] build their dam," and "[e]very summer that dam is torn up by the Government." Kipling warned visitors that "if you turn up a byway there is no more of you, and the beavers work your buggy into next year's dam."[41]

No doubt Kipling placed little stock in this image as profoundly descriptive of the war for permanence between the American Republic and the Republic of Nature. And yet the park experience did offer occasions for a close examination of American civil institutions and national character in their possessive relationship to the natural world. Elsewhere

in his Yellowstone account Kipling showed with candor a willingness to see in the American tourist an unflattering picture of excessive national pride, "the calm with which these folks gathered together and commenced to belaud their noble selves, their country, and their 'institootions' and everything else that was theirs."[42]

This brand of civic optimism as a form of wilderness socializing, while it seemed to Kipling's "bewildered ears" as "wild advertisement, gas, bunkum, blow, anything you please beyond the bounds of common sense," nonetheless struck deeply at the heart of Americans' anxieties about the lack of tradition and a cultured past, which made the nation appear inadequate in comparison with Europe.[43] Roderick Nash has argued the manner in which Americans seized upon their "natural environment" as "one of the few bases on which a favorable comparison could be made with other nations," and Barbara Novak has discerned in the fascination with geology an assertion that in the New World landscape lie a history both purer and older than anything to be found on the other side of the Atlantic.[44] Americans, then, could curb their fears of corruption and decay in a rejection of Old World history, even as they laid claim to geologic time as a sort of prologue to the real story, which, an understanding of providential history suggested was inherent in the New World landscape. The American continent, in other words, had lain fresh and unspoiled from time immemorial, only to insert itself into human history at the proper moment. European civilization had overwrought, and thus destroyed itself; yet oddly enough, the future in America held almost unlimited opportunity for fresh development and a human agency nearly divine in its sanction.[45]

Other European visitors to the park noted this urgent tendency to invest the landscape with nationalist historic significance. Nicholas Mohr, a German publisher visiting the region in 1883 by invitation of Northern Pacific Railroad financier Henry Villard, remarked that:

> The Natural wonders of the Yellowstone Park are a great attraction for all the Americans. America's sense of vanity is strengthened by having something besides Niagra Falls that can be found nowhere else in the world. Out of a sense of flattery, the people in the United States honor and respect this gift of nature and make pilgrimages to it as to a Mecca.[46]

Sidford Hamp, a young man from Great Britain, writing from Fort Hall, Idaho before traveling into Yellowstone with the Hayden Expedition of 1872, assessed the American anxiety about its lack of history from the perspective of schoolboy envy. His entry for July 4 states that "[t]his day is what the Americans call 'The Glorious Fourth' dash 'em! because they got their liberty on that day, a short time ago, (I don't know when)." He pined "[h]ow jolly it must be for the Americans to have no history to learn," but astutely observed as well "I suppose that is why they push ahead so, because they want a little."[47] There may well be that in Hamp's assertion which points to Americans' interest in bequeathing a history of accomplishment to their progeny; however, the tone of impatience implied in their "pushing ahead so" tends toward an ideology of national optimism which comes near suggesting that the geography of New World landscape was an arena in which Americans felt they might overtake their own future, as it were, in which the sense of prophecy might be so certain in the text of the civilizing project that the future could be regarded as history, though its realization remained a work of human effort.

That so much of Yellowstone itself lie essentially unredeemed mattered little, it seems; after all, Americans would find in the creation of

the park the means to proclaim at least an ambiguous sense of the magnitude and success of westward expansion. Teddy Roosevelt, for example, saw in the need to preserve Yellowstone a threat to the very democratic principles which had been hailed as the agents of progress and expansion. "Above all," he argued, "we should realize that the effort toward this end is essentially a democratic movement." This author of *The Winning of the West* registered his concern that private interests might withdraw areas like Yellowstone from their potential use for public recreation, though he assured his audience that it was "entirely in our power as a nation to preserve large tracts of wilderness... as playgrounds for the rich and poor alike." For Roosevelt, the consummate "believer in manliness, and therefore in manly sport," the wilderness experience seemed the last outpost in the maintenance of a rugged citizenry needed for the work of progress itself. His interests lay not only, or even primarily, with the park's landscape and natural features, but also with its population of game animals, which he felt should "continue to exist for the benefit of all lovers of nature, and to give reasonable opportunities for the exercise of the skill of the hunter, whether or not he is a man of means.... The wealthy man can get hunting anyhow," he argued, "but the man of small means is dependent solely upon wise and well-executed game laws for his enjoyment of the sturdy pleasures of the chase."[48] The strength of the nation's institutions was, then, intimately bound up with the idea of a character best cultivated in various facets of the wilderness experience, and Kipling's American tourists no doubt took pride not only in their landscape, but in their "noble selves" recreating en route to their virtue, and ultimately, in the service of their Republic.

Like Sidford Hamp, Kipling found Independence Day the quintessential occasion to take up a discussion of American attitudes. Unlike his young predecessor writing from a rude military post, however, he responded to what, by 1889, had become a highly evolved social scene evoking comparisons with European tourism, even though the National Hotel at Mammoth Hot Springs resembled "a huge yellow barn" by most accounts.[49] Kipling arrived in the park on the fourth, having "passed other carriages of trippers, who had done their appointed five days in the park," and who, he stated, "yelped at us fraternally as they disappeared in clouds of red dust."[50] Yellowstone was, of course, a destination for tourists from around the globe, and yet clearly he must have felt something of an outsider when, after the stages unloaded visitors on the veranda of the National, he noted with disdain that they "poured into that place with a joyful whoop, and, scarce washing the dust from themselves, began to celebrate the 4th of July."[51] What follows in Kipling's account is remarkable in its depiction of an impromptu performance by American clergy of a nationalist oratory whose traditions reached back to the Puritan Fathers.

"[S]itting on the landing of the first floor," Kipling notes, these ministers "rose up and told [the American tourists] they were the greatest, freest, sublimest, most chivalrous, and richest people on the face of the earth."[52] Sacvan Bercovitch has noted in this type of exuberant oration a long-standing, and peculiarly American form of narrative meant to secure consensus in the national vision, widely recognized since the publication of his work on this subject as the "American Jeremiad." Though its roots are properly linked to the religious and theocratic designs of the Massachusetts Bay settlement, Bercovitch's argument for the tenacity of

the genre carries the Jeremiad through a series of historical shifts which, while by no means fully secularizing the idea of national progress, do increasingly fasten upon critical moments in American history for a typology of national destiny.[53] The 1776 Revolution had provided later generations with a vast complex of rhetorical strategies for placing the young Republic squarely within an ongoing narrative of national mission, and which, when understood in the context of later days, essentially charged American citizens with the task of completing that which had been begun by the founding fathers through great sacrifice and bloodshed. Whether or not such oratory held secure to Protestant notions of Providence as the guarantee of continued success, American patriotism was significant enough in itself to make of the Fourth something of "a religious observance," as Rush Welter has described it, "in which the Declaration of Independence ... served in lieu of a religious text, and the oration in lieu of a sermon."[54] At any rate, Kipling's Yellowstone Tourists responded religiously to the clergy's reading of the Declaration of Independence and their claims of American greatness appropriately, for at the end of the "patriotic exercises" on the landing of the National Hotel, he noted" they all said Amen."[55]

The clergy in Kipling's account seem to have understood that the main text of national mission had moved westward onto the frontier, though they had little difficulty locating a sense of continuity with the Republic's legacy, or with its fundamental notion of a virtuous and unified citizenry. For after one of the ministers "bade the world note that the tourists included representatives of seven of the New England States," he turned to the value inherent in touring the Far West, suggesting even "that this running to and fro upon the earth, under the auspices of the

excellent Rayment, would draw America more closely together."[56] Admittedly, Kipling's journey into the park took place in the year directly preceding the momentous national census of 1890, which would cause Frederick Jackson Turner to proclaim that "at the end of a hundred years of life under the Constitution, the frontier has gone, and with its going has closed the first period of American history."[57] In this respect at least, it seems that tourists, of all people, might have rested content, enjoying the fruits of Westward progress. And yet this drawing together of the American citizenry remained a tenuous project in the eyes of the clergy, who, at any rate, felt the need to remind Westerners of "the perils that they of the East had surmounted by rail and river."[58] Tourists could indeed rejoice that the nation had carved its institutions out of the North American landscape; the precarious nature of American leisure and travel, however, served to remind them that much remained to be done.

To be sure, for all the paeans to the success of expansion echoing Turner's thesis, recreating in Yellowstone required a good deal of resolve right through the end of the Nineteenth-Century, and into the Twentieth. The Raymond and Whitcomb Travel agency had itself not seen fit to offer tours into the park until just previous to Kipling's trip, "feeling that even all its wonders would not compensate the travelers for discomforts he must inevitably encounter."[59] The system of roads to major attractions remained poor, stage drivers were notoriously reckless, hotels could be poorly built and overcrowded, and the weather has always been unpredictable. Americans, however, were adept and eager to experience the edifying effects of such hard won leisure. Herr Mohr, who held out hope for a tram system in Yellowstone to alleviate "the strain and sweat needed to pay a visit to this wonder of nature," noted with amazement

that "[t]he Americans have great patience and endurance. And even when it is only for pleasure, a thousand mile ride on the railroad is for them no insuperable problem."[60]

Nor was such a suffering, but willing figure a novelty born with the creation of national parks. Washington Irving, for example, who had worried that during his extended tour abroad his friends had thought him "beguiled by the pleasures and splendours of Europe, and 'treading the primrose path of dalliance,'" turned to the Oklahoma prairies as a fit landscape to lament that "[w]e send our youth abroad to grow luxurious and effeminate in Europe," and foreshadowing Roosevelt's demand for a rugged citizenry, he stated "it appears to me, that a previous tour on the Prairies would be more likely to produce that manliness, simplicity, and self-dependence, most in unison with our political institutions."[61] Within the symbolism of American wilderness recreation, then, lay also a call to work, and the reminder of implicit responsibilities to service as a citizen of a democratic republic.

This burgeoning image of the working vacation, however, remained curiously at odds with the Yellowstone landscape, which, in addition to its pervasive defamation as a realm of the infernal, was also inscribed with a concept of "play," which suggested an image of natural forces seemingly bent on mischief and counter-productivity. For Kipling, the geyser basins of the park were "given over to the sportings and spoutings of devils who threw mud and steam and dirt at each other with whoops and hallows and bellowing curses," and the reader can hardly miss the irony in his depiction of Hell's Half Acre as a "park-like place" where the guide "Tom suggested we should get out and play with the geysers."[62] It was, and has remained a commonplace to refer to an actively spouting geyser as

"playing," though rarely did visitors express a sense of ease in the presence of the sort of displays which were produced in the basins. Witnessing an eruption of the Giant Geyser in the Upper Basin, Reverend Stanley described it as "one of the grandest displays of water-works ever beheld by mortal eyes," yet he left little doubt as to the seriousness of the forces at work. "The fountains of the great deep seemed literally to have been broken up and turned loose again upon our sinful world," he stated. "At first it appeared to labor in raising the immense volume, which seemed loathe to start on its heavenward tour, but now it was with perfect ease that the stupendous column was held to its place, the water breaking into jets and returning in glittering showers to the basin." Characteristic of a host of descriptions which depicted the geysers as potential sources of productive energy, Stanley felt compelled to gauge the spectacle in terms of work rather than play, noting the "steam rushing upward from its chambers below, of which, if controlled, there is enough to run an engine of wonderful power."[63]

The forces at work generating geyser displays, however, often resolved themselves into delicate displays of light and color which, if imposing in scale, nevertheless furnished visitors with a sense of design in nature, and for many, a spiritual experience suggesting at least a momentary assurance of benevolence guiding the release of steam and superheated water. For Stanley, the actively spouting geyser seemed a sort of axis mundi at the heart of the world, which, as Mircea Eliade notes, is the image of "a universal pillar... which at once connects and supports heaven and earth and whose base is fixed in the world below."[64] There is, Stanley suggested, a fundamental purpose at work in the fearful underworld of Yellowstone. The labor of geothermal forces essentially

fashions an aesthetic release to the earth-bound tourist, whose fears that "Hades would feign disgorge itself and submerge our world" are dispersed in the "heaven ascending, graceful volume," the "steam rising higher, and still higher until lost in the heavens above."[65] Though he had "felt like baring his head and bowing the knee" before the sublime scenery of the Grand Canyon of the Yellowstone, Stanley intimated as well a spiritual/aesthetic vision more in line with the geysers when he expressed his hope that "if such slight glimpses into the ways and works of Nature's God are as manifest here below, obtained as they are with so much effort,... what must [the soul] enjoy when, freed from this cumbersome tenement, and with ... expanded and unobstructed vision" it can gaze freely upon the face of God, and "revel amid the untold glories of that 'temple not made with hands, eternal in the heavens."[66]

Stanley's vision of the Celestial City suggests a tenuous relationship with that work of human hands he witnessed while overlooking Bozeman and the Gallatin Valley, since it so squarely locates its eschatology outside the realm of human history. This is hardly to suggest, however, that hopes and concerns for temporal progress merely fled before wilderness revelations of the eternal. To be sure, nineteenth-century Americans were capable of thinking about the nation's future in exceedingly complex visions, ranging from a rich legacy of millennial rhetoric, to a classical republicanism, whose ideology of the civic state articulated the desire for permanence amidst fortune's vicissitudes.[67] It seems safe to assert, however, that progressive models of history essentially view the course of secular events as an attempt to get beyond the very problems of history as they have made themselves felt, a concern which Frederic Jameson suggests forms "the unity of a single great

collective story" whose "single fundamental theme" is "the collective struggle to wrest a realm of Freedom from a realm of Necessity."[68]

The Christian scheme of redemption, however, could lay hold of both providential history (with its promise of otherworldly salvation, and Freedom as the "end" of history), as well as an ideology of temporal progress, so long as the nation's millennial hopes could, by the middle of the century, be construed, as Tuveson argues, in terms of a sort of super-productive society, marked by the "incessant busyness (sic) of the people... engaged in 'industry,' in making inventions, in public service and the like," where the "command 'Work!' looms over the Happy Time."[69] Indeed, secular ideas of progress could be so purely bonded to the Christian redemptive scheme as to render the human harnessing of natural forces the underlying key to the full blessings of Providence. In 1898 for example, Methodist bishop Henry White Warren assured his audience that forces latent in natural processes were the medium through which to "secure the cooperative power" of God; it remained only a lack of insight and initiative on humankind's part, who remain "the slaves of ignorant circumstance." Warren argued:

> There is power of every kind everywhere in plenty. All the Niagras and Mississippis have run to waste since they began to thunder and flow. Greater power is in the wind everywhere. One can rake up enough electricity to turn all the wheels of a great city whenever he chooses to start his rake. The sky is full of Pentecosts. Power enough, but how shall we belt on? By fasting, prayer, and by willing to do the will of God.[70]

Warren, who devoted a chapter in his book to the Yellowstone geysers, admitted that the thermal features in the park were a "tremendous manifestation of uncontrollable forces." Yet elsewhere he stated too that it was "equally plain that there is intelligence, thought, and plan" behind the agencies of the natural world. "They are not blind

Samsons grinding in a prison- house," he argued, "and liable at any moment to bring down in utter ruin every pillar of the universe on which they can put their hands."[71] Though clearly the bishop may have thought twice about "belting on" to a geyser, the park itself was a lesson in the divine plan for utilitarian knowledge. And if he was willing to acknowledge that "[s]ome men never stand in the presence of an engine in full play without a feeling of reverence, as if they stood in the presence of God--- and they do," it seems equally certain that he saw in the landscape a redemptive schema offering more in the way of titanic productivity than in the complacency of pastoral bliss amongst the citizens of the Celestial City.[72]

Though Warren seemed content with revelation while gazing at Yellowstone's geysers, others were less comfortable with the expenditure of energy on so grand a scale. Charles Dudley Warner argued that "[s]team enough is wasted here to run all the Western railways," and even the naturalist John Burroughs stated "[o]ne disliked to see so much good steam and hot water going to waste; whole towns might be warmed by them, and big wheels made to go round."[73] Yet though other national attractions like Niagra Falls clearly evoked images of inexhaustible energy which, John Sears states, "made it an apt emblem for the resources of a new nation," Yellowstone's thermal activity could appear uncontrollable, excessive and erratic; indeed, to some extent, the general approbation of the Old Faithful Geyser as the park's best feature rested on its reliability. "Bless him!" stated Margaret Cruikshank, "He is so entirely all that we had anticipated... playing for us every hour, that we learned to love him. She noted that this attitude was "the general feeling, enhanced by the disappointment caused by the very capricious conduct of

his kinfolk."[74] Moreover, witnessing the displays from the basins, the Yellowstone tourist could hardly help but acknowledge a model of natural force which seemed bent on exhaustion rather than plenitude; the labor of nature appeared problematic in its endless cyclicality, its expenditure of force followed by a calm which only seemed to renew its capacity for exertion. Owen Wister found this aspect of Yellowstone touring so compelling he intimated that successful geyser gazing held fast to the tenets of the workplace, and his image of American leisure is alarming in its reproduction of the daily grind from which tourists had sought relief in the first place. Aided by the mischief of a young employee at the Upper Geyser Basin, Wister takes this form of leisure to task:

> We would be sitting tilted back, reading our mail, the tourists would have ceased talking and be lounging drowsily, the boy would be at the door, motionless as a set steel trap. Suddenly the trap would spring, the boy would catapult into the door, and in his piping treble scream out:
> "Beehive's a-goin' off!"
> At which every tourist instantly started from his chair, and a leaping crowd gushed out of the hotel and sprinted down over to the formation to catch the Beehive at it. Beehive finally quiescent, they returned slowly, sank into chairs and exhausted silence; you could have heard a mosquito. But the steel trap was again set, sprang soon, and again the silence was pierced:
> "There goes Old Faithful!"
> Up and out they flew once more, watched Old Faithful, and came back to their chairs and to silence more exhausted.
> Was the boy exhausted? Never. It might be the Castle, it might be the Grotto--- whatever it might be that pre-Ritz-Carlton bell-hop routed those torpid tourists from their repose to set them trooping across the formation to gape at some geyser in action, and again to seek their chairs, feebler each time. Has he in his mature years ever known more joy? I doubt it.[75]

Admittedly, the majority of Yellowstone tourists of the day who stayed at hotels had less to be concerned about with regard to the problems of getting a living than those tourists known as "sagebrushers," who fashioned for Caroline Hughes the sweeping "farmer class," visitors

who in their own conveyances spent their vacations "driving through the Park, camping out at night for a pleasure trip." If she had felt harried by her guides in her visit to the geyser regions, she seemed pleased at least with two girls who "had collected ferns and flowers and looked thoroughly happy," and who in this manner "have a thoroughly restful, cheap holiday."[76] Yet while the ideology of leisure suggested to the upper class the mark of favor by which to gauge their freedom from the rounds of toil, the processes at work in Yellowstone's landscape seemed matched to the specter of Hannah Arendt's *animal laborans*, whose condition in the realm of human affairs constitutes the physiological imperative of life itself, and whose futility is traced in the fact that "the mark of all laboring leaves nothing behind, that the result of its effort [merely the capacity to labor again] is almost as quickly consumed as the effort is spent."[77] Arendt distinguishes "labor" from the activity of "work," the latter growing out of the surplus capacity of labor to get beyond the needs of its own reproduction, to fashion a world of enduring objects and comfortable surroundings. Beyond work lies the life of free contemplation, and insofar as this terminology itself fashions a redemptive scheme for the human life, the moral problem of effort rests in a "[c]ontempt for laboring, originally rising out of a passionate striving for freedom from necessity."[78]

Wister's account certainly hints at the labor inherent in geyser watching as ostensible leisure, though it ought to be the act of contemplation par excellence. In other words, contemplating Yellowstone's geysers may have indeed fashioned a sort of objectifying gaze creating a proper distance between the tourist and nature's toil. Yet Yellowstone nomenclature, and representations of the park experience remain striking in their blend of the profane and the toilsome, remarkably

so as relates to women's labor and domestic chores. Features like the Devil's Kitchen, the Devil's Frying Pan, and hot springs like the Handkerchief Pool, where Lee Whittlesey notes "[b]y 1906 guides were telling visitors that this was where the Devil took in washing," all presented aspects of domestic labor as cursed, and there is no real reason to think that such views were limited to the park, especially in women's conversations.[79] For it could be argued that the ideology of labor had, by the middle of the nineteenth-century, adjusted its attitudes to a society which had largely been freed from concerns with sheer subsistence, but which remained pejorative toward aspects of effort tied most closely to the life process. Bishop Warren, after all, understood the moral imperative at the heart of gravity as primarily bound up with male connections to housework. "I have seen in many lands men bringing to their houses water from the hills in heavy stone jars," he lamented, even though the "good Father has offered his all-power on all occasions to all men." What such blessings held out for women seems less certain, though one senses in the inherent design a pervasive sense of well-being, the model family having come to the revelation that "[g]ravitation was meant to do that work [transporting water], and to make it leap and laugh with pearly spray in every woman's kitchen."[80]

It is hardly surprising that thermal features throughout Yellowstone invited comparisons with kitchen chores, or with the laundering of clothes; yet the extent to which tourists took to these activities as a form of play reveals much about cultural attitudes toward labor. Moreover, while cooking or performing laundry in hot springs suggested pragmatism in a region where such services were hard to come by, visitors to Yellowstone found countless occasions to note the novelty of

taking nature to task in this manner; indeed, this fascination with wilderness domesticity forms one of the strongest unifying elements in a body of literature remarkably diverse in its origins. On the small scale, at least, tourists seemed to agree with Bishop Warren that nature held out its blessings for the weary, though as often as not its more capricious moods held sway in the park. The Reverend Stanley, for example, was amused by one "curiosity-loving, labor saving philosopher" who, "having a natural abhorrence of dish-washing" attempted the task in a hot spring pool. The account is a model seduction which, if playful, nevertheless structures a nightmare of the domestic fantasy gone bad; though the pool is "a lovely little basin," the philosopher soon discovers he has been, after all, "'taken in' for one time."[81]

As if rediscovering the logic of the marriage contract like some sort of Ur-bachelor, Stanley's philosopher sits "in his musings" before his dishes, until, "enticed by the bubbling of the water" he hits upon the idea to "give them over to his new servant,... reposing all confidence" in the spring as a wilderness helpmeet. Though Stanley gives no mention of the location of this diminutive spring, we learn of its christening as the "newly found Bridget," and after having handed over the task of washing, we find our bachelor "lost again in meditation upon the advantages of a life in Geyserland."[82] The scene hints at the pastoral, and it could be argued that the momentary sense of order and repose is related to what Carol Merchant has suggested is the ideology of a feminine nature "tamed and subdued... to enhance the comfort and soothe the anxieties of men distraught by the demands of the urban world and the stresses of the marketplace."[83] Once the union is joined, however, Bridget shows herself the shrew, and "the hitherto quiet little spring, as if indignant at the

insult offered" is soon "writhing and spouting and splashing about at a fearful rate, and tossing knives and forks, spoons and plates, and cups, around in utter disgust."[84] Though John Sears has argued that nineteenth-century tourism "was never gender-identified," clearly gender constructs influenced representations of the park, and ideologies of the feminine as both subdued and potentially chaotic shaped ideas of Yellowstone's landscape as they had other regions of the American wilderness.[85]

At moments, Yellowstone's landscape presented to visitors' imaginations a kitchen scene comic in its proportions, yet no less scrutinized from within prevailing values of domestic economy. Visiting the odd formations of the Biscuit Basin area in the farther reaches of the Upper Geyser Basin, for example, Dorothy Pardo quipped "[i]f these were biscuits, they must have been the effort of a 'Mrs. Newlywed,' for we walked on them over to the rim of Sapphire Pool."[86] Cooking food in Yellowstone's hot springs was the most obvious manner in which to take advantage of natural forces at work in the park; to be sure, Jim Bridger's stories of cooking fish on the line seemed borne out at the Fishing Cone hot spring in Yellowstone Lake, where tourists were often photographed in chef's garb, modeling in a sort of subsistence fantasy tableau.[87] To Margaret Allen, the sight of "neighbors" doing laundry in the springs reminded her that "we have a ham in our wagon," and spending little time wondering "why should not that be cooked in the same way," she states "[t]he Devil's Well is near, and soon our ham, in a strong sack fastened to a pole, is cheerfully bubbling away." The risks were made clear, however; the Fishing Cone could erupt to heights of forty feet, as it did occasionally in 1919, and Allen had to admit that "[o]ur potatoes," cooking after the

fashion of the ham, "are not so successful, for our bag breaks, and down they go to whoever the owner of the well may be, for a perpetual potato-soup."[88]

Losses of the sort the Allen family incurred may have been common enough in the Park, yet nowhere did tourists note them with such a blend of amusement and dismay as they did with regard to attempting laundry in springs and geysers, far and away the most popular form of domestic tourism in Yellowstone. Tourists were aware of the fact that pouring soap into a geyser could cause it to erupt, and they reveled in Park lore which held that a Chinese launderer who, in one account, "pitched his tent over the spring, thinking to save his fuel bill in the cause of heating the water.... All went well," stated Joe Chapple, "until one day he accidentally dropped into the boiling pool his big cake of laundry soap." What followed was a denouement which was to amuse tourists into the twentieth-century who held typical attitudes with regard to race, Chapple's version gleefully stating "Blu-u-s-h! up went this hot spring into a geyser, taking with it the Chink and his tent," while "[n]ever since that day has even a remnant of that Oriental been found."[89]

Such stories hardly deterred Yellowstone tourists from their own attempts at geyser laundering, however, and most seemed to agree with Reverend Stanley that here nature brought out the best in the spirit of human ingenuity. Stanley hinted that in Yellowstone, the urge for novelty often blended with practical ends, and he noted that as "curiosity-loving mortals" are '[a]lways seeking something new... . the geysers afford opportunity for more than one inventive genius to try some unknown and hitherto unthought-of experiment." His genius wonders "'Can we not avoid the labor of laundry-work?'" and the question is no sooner posed

than he "immediately commits his task to Old Faithful, with orders to return his washing faithfully in one hour."[90] Yet perhaps because nowhere did nature show itself more unreliable than in this area of domestic labor ("many articles that go into her chambers 'ne'er come out again'"), geyser laundering was often seen as an enterprise both futile and costly, though one which seemed to tempt tourists to their strongest literary efforts. Most remarkable, perhaps, is the delightful mock-heroic "The Tail of a Shirt" written by Sidford Hamp and friend C. Spencer, a poem which plays upon the trivial much in the manner of Pope's "The Rape of the Lock." Borrowing on the epic device of the hero's journey to the underworld, "The Tail of a Shirt" anticipates in 1872 what was to become a commonplace fascination with features like the Handkerchief Pool, which would take an article of clothing into its interior and return it "clean" by the normal circulation of its water.[91]

Sending clothing down a geyser seemed to construct a surrogate figure fit to experience the Park's nether regions, especially for those who may have shared Charles Warner's opinion that "[b]ewildering as all this spectacle is to one's idea of a normal and orderly world, I was more impressed by what I could not see.... It is what is underneath this thin crust, it is the state of things underground that appeals to the imagination."[92] Margaret Cruikshank, visiting the Park at the same time as President Arthur in 1883, wrote of a "funny scene" related to her by a Mrs. Gobeen in which clothing placed by Arthur's escort into Old Faithful was rudely ejected, "filled out with steam to superhuman proportions," so that "for a while it seemed as if the air was full of swollen fragments of men."[93] "The Tail of a Shirt" inscribes its hero shirt with the capacity to produce an account of the journey, ("From that region vile, and coaly/ It

returns a sight more wholy ") yet the overall effect of the poem is to diminish through its satire natural forces whose potency was only too evident in the Yellowstone landscape.[94] And while the outcome betrays a sense of concern that the underworld make itself too well felt at the surface where humans conduct their mundane affairs, the satire nevertheless attempts to diminish and contain the potential for chaos, and the underlying motives of harnessing Yellowstone's energy remain intact, if only in the form of Spencer's "inward hope/ Of washing out his shirt with soap."

"I washed some clothes today in a hot spring," Hamp noted in his diary entry for August 16, 1872, "by simply boiling them for an hour." He noted another washing on the 17th which was similarly uneventful, yet on the 20th, a layover day for the Hayden party, he stated that "Spencer and I made a 'Poem' on the adventures of one of the formers (sic) flannell shirts," which had, on the washing of the 17th "went right down" a hot spring, though it "3 hours afterwards came up qu[i]te clean, a proof that the 'Devil is not so black as he's painted."[95] "The Tail of a Shirt," then, is the account of the shirt's journey to the underworld of the Yellowstone landscape, a Stygian world "where the Geyser waters mix:"

Where waters dash
With roar and crash
O'er blackened stones,
And Dead men's bones,
Where Cerebrus yells,
Where sulphur smells,
Where vaults resound
With horrid sound,
Where crawling things
All legs and wings,
Sport in the mud,
On the banks of the flood.

This is arguably an apt description for an active thermal region not fully understood to this day.[96] Yet though the poem makes much of "the wonder what can be the source/ Of such tremendous hidden force," Spencer's needs are modest:

> And when the article was lathered,
> His shirt into a ball was gathered.
> And thrown into the spring to boil,
> To save himself from needless toil.

Like Reverend Stanley's labor-saving philosopher, Spencer too turns to the contemplative life: yet his thoughts hardly stray from the task at hand, which seems to have taken on a pleasing nature of its own under the gaze of the young man, who: "as he saw the dirt was leaving,"

> Which to his shirt tails had been cleaving
> He heaved a sign of inward peace,
> That he'd washed out those signs of grease
> That came of sinker-bread and bacon
> And things which tend to stomach achen

There is the intimation in these lines that camp life is not, after all, a flawless one, and it should be remembered that the 1872 Hayden Survey mapped some 9,000 square miles of little known territory, were out from early July until October, and though traveling through a newly established national park, had no amenities to speak of.[97] Moreover, Spencer finds his reveries broken by the disappearance of the shirt down the mouth of the spring, and the day of rest is shattered in a frenzy of activity that is painful to watch:

> How like a raving maniac
> He strives to get his garment back,
> But all his strenuous efforts fail
> To grasp its fast receding tail.
> With open mouth his friend Hamp gazes
> His nostril spreads, his hair it raises
> Like porcupine or hedgehog bristles.
> His fingers spread, his eyes stick out,
> To see friend Spencer rush about,

In search of sticks, and crooked poles,
To poke about in hellish holes;
Till finding labour unavailing,
He throws them down and walks off wailing.

The scene certainly yields heroic effort on behalf of a domestic project; the "hero shirt," however, is the passive envoy caught up in the play of Yellowstone's forces, and though conducted through the dangers of the underworld "unscathed and unhurt," its "passage to the upper air" is trying if successful:

Its back "dog gone it" receives a tear.
And its tail gets into a deuce of a curl
As it's caught by the steam in its upward whirl.
With a twist and a turn, it is borne to the top.
And out of the basin it flies with a pop.
Then it descends a few moments later,
And hangs by one arm on the edge of the crater.

The 'tail' of the shirt is a text, then, of the mischief wreaked on human effort by natural forces in no wise to be trusted. And while the poem makes light of its moral by stating:

Friends if youve (sic) a flannel shirt
That you'd not have come to hurt,
Read this moral and be wiser,
"Don't wash clothing in a geyser,"

the fact remains that Yellowstone would, in years to come, play host to a number of tragedies as a result of natural phenomena, some of which, like the 1959 earthquake on the Park's western boundary, could alter the landscape so profoundly that the loss of human life seemed something of a sideshow.[98] Yellowstone tourists recreated at their own risk in the midst of forces constantly remaking the landscape, and most visitors retained a sense of awareness that their leisure never fully removed them from the concerns of their daily lives. Mrs. Corthell, for example, appeared

exasperated at moments with the Park's frightening underworld, and she argued that it was, after all, "a pokerish kind of pleasure trying to enjoy the ravings of the demons from the bottomless pit" at Yellowstone Lake's West Thumb. Her problem was, as it has ever been, watching out for her progeny, and she complained that "[a]s for me I was kept busy counting the children. Every time one moved I felt certain he would stumble into a boiling, walloping vat of mud." Laying waste to descriptive conventions which had shaped countless promotional and guide books, she suggested further that the Park's scenic wonders were small consolation to the woman who found her cares had traveled with her to Wonderland, for she made it clear that though "it was delicate rose, emerald green, or heavenly blue mud did not reassure me.... Playing with silver-tipped waves or silver-tipped bears was safe in comparison."[59]

Yet if visitors like Mrs. Corthell were content to put distance between themselves and Yellowstone's thermal phenomena, others seemed vaguely disconcerted that nature should not be held more to account in the Park, and many intimated that a moral issue was at stake here in the antechambers of the American landscape. Kipling, who had traveled through Yellowstone in the company of an elderly couple from Chicago, noted that "the old lady" was prone to see in the Norris Basin a lesson for backsliding sinners "that the Lord has ordained a Hell for such as disbelieve his gracious works." "If," she asserted, "if we find a thing so dreffel as all that steam and sulphur allowed on the face of the earth, mustn't we believe that there is something ten-thousand times more terrible below prepared untoe out destruction?" Her husband, however, if he had religious views at all, might have argued with Bishop Warren that Americans were ignoring the blessings of Providence by allowing nature to

run riot in Yellowstone, and he capped her ejaculations of "'Good Lord!' at thirty second intervals," with commentary on the "'drefful waste of steam-power," which he simply stated was a "wicked waste," by the time they reached Hell's Half-acre.[100] It would be a few year's before the Chicago couple would have the opportunity to gaze upon the channeled energy of the forty-odd steam engines in operation in Machinery Hall at the 1893 World's Columbian Exposition, where, if they sought with Bishop Warren the assurance that America could "belt on" to the latent forces of nature, they would find it in the "miles of shafting, pulleys and belting connected with the thousands of exhibits that were in motion every day during the Fair."[101] Though a writer for Harper's Weekly suggested that most of the Fair's power needs were supplied by electric wiring, and that with regard to belts and shafts, "[n]o one would now think of taking power for any considerable distance by such wasteful methods," the source of all this energy retained a similarity to the Yellowstone landscape. "In front of the boilers is a gallery for spectators," the writer noted, so that visitors can conveniently inspect the system. Probably visitors to this power-house will afterwards readjust their notions of the infernal regions, for the burning oil will make a dreadful roar, and the light will make the firemen seem uncanny beings."[102]

For the moment, however, Kipling's party had to remain content with a token washing at Larry's if they sought evidence of "man's conquering rule over earth, fire, and flood," hailed as the Columbian Exposition's true achievement.[103] Larry, whose sense of practicality allowed him to issue his statements with authority, told his diners "'you'll all go an' wash your pocket-handkerchiefs in that beautiful hot spring round the corner,'" where he noted "'ther's soap an' a washboard ready, an'

'tis not every day that ye can get hot water for nothing.'" And Although Kipling suggested that "Larry alone" knew "[w]hy we should have washed pocket-handkerchiefs," he conjectured that "[i]t appeared to be some sort of religious rite."[104] Mircea Eliade speaks of the difference between ritual labor and "similar operations performed outside of the sacred period," suggesting that ritual structures a mythic moment to house the acquisition of skills or technology in the group consciousness. The Prometheus narrative is familiar enough in this respect, while Eliade offers the example that a "boat is repaired ceremonially not because it is in need of repair but because, in illo tempore, the gods showed men how to repair boats."[105] No one would have accused Larry of offering his clientele holiday fare; yet there was that in his services which made sacraments of hot water and soap.

There is little need to strain the interpretive context here; yet to suggest that tourism in America structured its temporal experience as a "ceremonial time" outside of daily affairs could hardly be argued from within the history of Yellowstone accounts or its promotional literature.[105] Moreover, the images of ritual labor in Park accounts are simply too pervasive to be viewed outside of their ideological content for the culture at large. That few tourists took these forms of playful work seriously is clear enough; that they recorded them with a remarkable consistency suggests that they understood the extent to which their attitudes toward leisure were shaped within the context of labor. In this sense, tourists at play with models of domestic labor in Yellowstone could exorcise the demands placed upon their productive energies, while constructing at the same time a collective memory of the harnessing of natural forces in illo tempore, the initial gesture in the civilizing project. The distance between

the Mammoth Hot Springs of the Park and the mammoth power plants in operation at the Columbian Exposition, then, comes to constitute the space of a historical process in the face of which nineteenth-century Americans exhibited an amazing optimism. So long as they could feel secure in the extent to which the national narrative was one of "belting on" to the forces so visible in the Yellowstone landscape, tourists would continue to recreate in the midst of the inferno; if the Park itself lay unregenerate, if all of nature seemed caught up in the endless cycles of labor so evident in the geyser basins, few nineteenth-century visitors would have felt that they shared such a fate.

Creators of promotional material for Yellowstone understood this clearly enough. In 1904, for example, the Northern Pacific Railroad marketed its Wonderland guide with a lurid cover which, when spread open, depicts a Yellowstone landscape in which the foreground features stripped and fallen timber giving way to a background apparently lit by flames and cast against a scarlet sky. The right half of the picture shows a portion of an orb (though not clearly either sun or moon) obscured by the steam rising from a geyser basin and features the Wonderland title in a font seemingly constructed of lightning bolts. The rock formations of the scene are decidedly upright and angular; there is not, in the whole scene, a spot where it could be conjectured that a human figure might safely stand. The cover illustration is remarkable in its open hostility, though the author Olin Wheeler noted inside that:"

> If, as has been said, this is the safety valve of the globe, or any considerable part of it, it may be a very safe place to be, for steam boilers seldom, perhaps, burst near the point where the safety valve is attached. Certainly no Mount Pelee or Vesuvius has broken forth here in historical time, , notwithstanding the boiling and eruptions that... have... been going on for nearly a century. The record shows that the operation of the thermal forces in Yellowstone Park has been

> remarkably uniform, notwithstanding the attempts of a few individuals to prove the contrary.[107]

Wheeler encouraged potential visitors to Yellowstone by suggesting that thermal forces operated within nature's regulation of its own machinery. Yet it should be noted as well that the 1904 season saw the opening of the Old Faithful Inn at the Upper Geyser Basin, and it was in the Park's developing hostelry and service operations that tourists could be most assured that Yellowstone featured human comfort amidst this reckless display of force. Indeed, Old Faithful Inn seemed to provide some of the best evidence to date that homo faber, the toolmaker, and builder of things permanent had come of age in the wilderness.[108] Though Anne Farrar Hyde argues that the Inn "appeared as if it had been spawned by the same forces of nature that had created Yellowstone," and that "the gnarled and twisted branches used as exterior trim and porch railing demonstrated the same malevolent forces that had tortured the earth in the thermal basins," she notes as well the achievements of workmanship and the architectural feats which were showcased in the design.[109] And while the intent of architect Robert Reamer was to build a structure in harmony with its surroundings, the Inn was, after all, an equivocal "product of the forest;" clearly given form and form-giving, the hotel "built with ax, saw, and hammer," did not "give of bird or bush" like nothing else in Yellowstone.[110] From the balcony, tourists could sit in the evenings and gaze at that most orderly of the park's geysers, framed in the beam of a searchlight mounted on the roof. They could retire to comfortable rooms while the world of the Upper Basin kept to its endless rounds. They could relax in hot baths while the springs scalded the landscape. And finally,

they could utter their prayers in beds undisturbed by the shocks and tremors of forces wreaking havoc in the nation's playground.

Tourists remained adept, however, at finding rhetorical strategies which made light of natural processes in the park. Most seized in some fashion on Yellowstone's history as an arena of the profane, which to some suggested the specter of hard labor performed in mythic proportions. "Old Vulcan seemed hard at work with a mighty sledge hammer far down in the bowels of the earth," noted Wilbur Sanders in his journal of an 1881 trip to the park, though the quip left him uneasy, and he admitted that "[i]t was a queer sensation to know what a terrific power was being exerted such a short distance, as it seemed, beneath our feet."[111] Others appeared unwilling to brook the notion that the park's underworld housed any sort of productive efforts whatsoever, and held to the tenet that the park was an arena where natural forces defined themselves through mischief and unrestrained play. Joe Chapple, who imagined Yellowstone's nighttime landscapes an "Inferno, illuminated by the lurid gleam of phosphorous vapors," suggested that the park had its moments of "high carnival in ghoulish glee," and in what was surely one of the most strained metaphors in the history of Yellowstone accounts, tested his readers with the "Midnight Revels on the Devil's Golf Course."[112]

The chapter is remarkable primarily in its blend of wanton force with a landscape being created and shaped in a series of cosmic blunders, so to speak. Chapple's fantasy seems to squarely challenge the Cartesian argument by design, and though it depicts the Devil's recreation as an effort driven by malevolence and a sense of ultimate failure, the activity itself clearly is bound up with the making of the American landscape; the colossal imagery itself trains the reader's perspective onto broad national

vistas, and the links are set expansively "a'top o' the Rockies." "[F]rom a tee on the heights of the Teton mountains," however, Chapple brings the Devil's game securely into the boundaries of Yellowstone, where, "mopping the perspiration from his brow... forming rivulets that flow down the mountainside," Satan nightly recreates in, and re-creates the face of, a significant piece of the national landscape.[113]

Yet Chapple would hardly have suggested that the "nocturnal golf game of his Satanic Majesty" shaped a problematic context for the drama of national progress. Yellowstone, after all, retained elements of sublimity that had to be negotiated in the round, and Satan finds himself at Inspiration Point overlooking the falls "entranced by the beauty of the spot," though he "forces himself to go on with the game." Perhaps no spot in Yellowstone fashioned a more consistent set of responses that reflected the assertion that God's presence was manifest in the face of nature, and the park's most appealing features clearly mar the game by evoking in Satan a sense of separation from the sacred landscape. As the course moves onto the Mammoth Hot Springs, he finds that "under the halo of Angel Terrace of snow-white purity, scintillating, mirror-like in all its pristine beauty," lie the "bunkers and hazards that must ever keep him from winning the contest against the heavenly battlements," and the realization of this fact becomes the fundamental force shaping the more profane features of the park. The endless re-creation of the face of the landscape in Yellowstone becomes the stage for a mock tragedy; in essence, the play rehearses Satan's role in the past and future of the redemptive scheme, a "flub score" in which "again tradition records the fall of Lucifer." For Chapple, whose Yellowstone account traces the moods of the park through the trope of an exemplary, and supratemporal day,

the Moonlight Revels revive nightly the biblical narrative. For those who held to the idea of America's role in the drama of redemption, the infernal imagery offered comic assurance that Satan was firmly bound in the game.

Chapple's "picture of Yellowstone Park as... a fanciful golf course dotted with brimstone greens and... sulphur hazards" ultimately yielded to the reclaiming project of an aesthetic nationalism; so long as the play of natural forces seemed restrained by more sublime (and more durable) elements in the park, Yellowstone itself was the model of American ascendancy over an unregenerate landscape.[114] If, in other words, the Organic Act of 1872 suggested that the country housed the landscape of original sin, American tourists with a broader vision could see beyond to its ultimate redemption, since, as William Helmuth rhymed in a poem "The Yellowstone Park and How it Was Named:"

> Yankees can sometimes, without doing evil,
> O'ermatch in sagacity even the Devil.
> For not long ago Uncle Sam came that way
> And said to himself, "Here's the Devil to pay.
> Successful I've been in all previous wars;
> Now Satan shall bow to the Stripes and the Stars.
> This property's mine, and I hold it in fee;
> And all of the earth shall its majesty see.[115]

Although this grotesque satire which shapes so much of Yellowstone imagery tends to understate complex attitudes toward nature inherent in park accounts, there is no doubt it was directed at forces betraying a rawness of power nearly illimitable to the human sensibility. Yet tourists continued, throughout the early history of the park, to make play of natural forces by resorting to infernal imagery . Visitors responded with zeal to the tradition of Satanic nomenclature begun by those, who in Stanley's opinion, saw in every "oddly formed rock or

mound, every river, gorge or glen... a fancied likeness to the hideous person whom they had been taught from infancy to dread, and the regions they had been warned to shun."[116] Though it is unlikely that even the earliest travelers to Yellowstone were intent upon an authentically profane lexicon of park features, those who were more properly "touring" the basins in later years would seize upon the infernal much more squarely as a mode of playful behavior which may have allowed them to feel secure in the park. Visitors seemed rapt with the idea of leaving inscriptions at some of the park's more infamous features; Mrs. Corthell noted one such instance at the Devil's Frying Pan in Norris Basin where she commented on a "guide board having several notes in pencil beneath... addressed to his Satanic Majesty." She recorded one in particular which went so far as to flaunt the prospects of hellfire as a jest, though the author spoke for a group who maintained they "were right glad to find [the Devil] out, whooping it up for the other sinners." Yet these tourists seemed fully prepared for retribution, and the author entered a plea by requesting that "when it comes our turn, please, dear Devil, we don't want to boil in your cave or bake in your kitchen or sizzle in your frying pan--- give us the dynamite route--- sudden and not too hot."[117] Free and easy concourse with the Devil fashions here a remarkable context for nineteenth-century leisure, and tourists to this day continue to find ways to articulate a brash sense of security in Yellowstone by tempting fate, a fact which has caused endless amazement and exasperation on the part of park rangers and personnel.[118]

Not everyone shared the sense of play that underlie infernal imagery in the park, however. Caroline Hughes, visiting Yellowstone in 1891 noted in her diary entry for September 10 a road "almost impassible

at one part where the pitch was very steep," and stated that "someone gave it the name of the Devil's Staircase and wrote it up on a big label." She seemed unconcerned, but remarked that "Mr. Hague had it suppressed at once."[119] Arnold Hague, a scientist with the United States Geological Survey in charge of the survey in Yellowstone, disliked the history of Satanic Nomenclature in the park; in a letter written to a professor R.B. Moore of Butler College discussing a paper written by Moore, Hague said:

> I noticed one name you used is Hellbroth Spring. I do not recall this spring and it cannot be on our maps, but it may possibly be one of those named by the Hayden Survey and omitted on our recent atlas sheets. I want to eliminate as much as possible Hellfire and the lower regions. We have succeeded in most cases, but in one or two like Hell Roaring Creek, which has fastened itself on the community, it was deemed best not to make the change.[120]

Hague doesn't state the exact philosophy guiding his attempts, yet bad press for Yellowstone on its own terms could hardly have caused a concern insofar as the park seemed to thrive on its reputation. More likely, he found it unseemly for a nation to treat its religious values so lightly. However, the fact that visitors to the park felt comfortable toying with the Judeo-Christian drama suggested, perhaps, the extent to which the redemptive scheme had been assured in the creation of Yellowstone as "the playground of man." And while one is never quite certain whether God had reclaimed Yellowstone for the Republic, or whether the Republic had reclaimed it for God, what is certain is that tourists to the park increasingly expected that their visits lend them a credible wilderness experience singed by raw natural force, yet replete with the blessings that civilization had laid down in shaping the artifact of a national park.[121] What made the progress of leisure in Yellowstone so unique was the manner in which it continued to confront the landscape of natural force

with the landscape of holiday. Park diaries and journals record a litany of aesthetic, contemplative responses to the visible exertion of Yellowstone's natural phenomena, and in so doing, tourists found ways to assess the national narrative from within the ideology of leisure.[122] It indeed seemed that natural forces labored in the park to no better purpose than the sheer pleasure of an elect national audience, a tour de force redemption of the wilderness, and a bold assertion of American ascendancy over resources which were as much a part of the landscape as the scenery. If, then, tourists like Chapple felt free to view the park's thermal areas as a review of the historical record, arguing that "[f]rom these geysers the onlooker learns the same lesson learned by Watt when he observed the boiling of his mother's teakettle" --- if, as he suggested further, "[t]he approach of the triumph of the age of steam is found in these hot springs which cover the largest area in the world," they could rest assured that they stood at the far edge of progress, and that the Yellowstone gaze was truly retrospective.[123] Finally, if they harbored fears that the incipient power seething beneath the crust of the park remained beyond the scope of human ingenuity and control, the 1904 Wonderland, at least, bade them note on their journey "The Lignite Coal Fields of North Dakota" which, after all, were "practically free from sulphur," and "[f]ortunately too... easily exploited." Moreover, if the "wicked waste" carried out daily in the park's basins raised the moral issue of unexploited resources, Wheeler suggested that, where properly managed, "[l]and, water, and fuel are the trinity that determine the building up of the waste places," and that if Yellowstone remained an unregenerate landscape, "North Dakota has them and they can be obtained for reasonable prices."[124]

Wheeler's claim appears to have been simply a reassertion of the notion of the providential landscape making its entry into history, as it were, though it is clearly a secular model of progress that dictates the imagery of *Wonderland*. And yet the vastness of the North Dakota fields could hardly have seemed to the nineteenth-century mind the construct of chaotic forces, even less likely the product of fortuitous circumstance in "the deposition of fine sediment on the mass of fallen and melted vegetation... and these conditions recurring again and again," producing "the many seams, or beds found at different elevations."[125] To be sure, in marked distinction to the reckless expenditure so evident in the park's geothermal activity, the lignite fields exhibited the patient storehousing of forces whose potential could be released just as patiently in a controlled, and productive process. The geological record suggested that from time immemorial, endless generation and consumption had been the model of nature's labor in the park; the prehistory of the North Dakota trinity, on the other hand, was replete with the logic of primitive accumulation in keeping with the genesis of a national economy.

Reid Badger has suggested that Americans had, by the latter half of the nineteenth century, begun to formulate the rhetoric of a "natural history of industry" in the face of a fluctuating and uncertain economy.[126] That the natural history of industry might reveal itself in the industry of natural history is simply the logical proposition in this context, though the rise of an urban-industrial landscape appears to have tempered somewhat the exuberance Americans may have felt in the midst of expanding productivity, as well as their willingness to project these attitudes as positive aspects of the natural landscape. The landscape of Yellowstone, however, could hardly have escaped an association with the industrial

complex; Owen Wister spoke for a host of visitors when he calmly observed of the Upper Geyser Basin that "[i]t might have been Lowell." He went further by stating that "the prospect suggested to my modern mind a manufacturing center in full swing," though he thought modernity a rather timeless concept and registered little "wonder those shooting columns scared the Indians of old."[127] Wister's uncertain response to the topography of industrialism and the conditions that went with it may have been of a piece with the mood of the culture at large. Certainly Americans had been exposed to an iconology of American manufacture in which the infernal shaped the moral climate of labor; Rebecca Harding Davis' Life in the Iron Mills was exemplary in its depiction of the Wheeling Iron works as "a city of fires:"

> Fire in every horrible form: pits of flame waving in the wind; liquid metal-flames writhing in tortuous streams through the sand; wide caldrons filled with boiling fire, over which bent ghastly wretches stirring the strange brewing.... It was like a street in Hell.[128]

Yet while tourists found numerous occasions to respond in a similar fashion to the scenery which met them in the geyser basins, most noted the odd juxtaposition of wilderness and industry as a complex relationship that was puzzling in its implications. Mrs. Corthell, for example, implied that the landscape surrounding the Midway Geyser Basin lent a sedate atmosphere to the "weird, ghostly spectacle" in which, she stated, "[a]pparently smoke stacks and steam engines are sending their cloudy columns above the dark foliage in all directions." It was the pervasive sense of quiet, however, which allowed her to maintain the contemplative stance appropriate to the tourist, and she checked the implications of her vision with the assurance that here there were "no cities, no factories, only the silence of the forest."[129]

John Sears has suggested that tourists in the park were, in general, not "bothered by the way Yellowstone's natural phenomena mirrored the machinery of the industrial age," and that they "saw no contradiction between their enjoyment of a bit of wilderness adventure and their faith in progress and economic development."[130] Of course, Yellowstone's early clientele was made up in the main of those who likely benefited in some way from industrialization; few of Harding Davis' "ghastly wretches" would have been found on the Pullman cars unloading passengers at the Northern Pacific terminal. To be sure, in their daily experiences, Yellowstone's tourists were probably no less adept at placing distance between themselves and the urban-industrial landscape than they were in constructing an aesthetic, contemplative space between themselves and the industrial wilderness of the park. Americans' fascination with technology in the late nineteenth-century was itself bound up with the consumption of images far removed from the context in which it actually functioned in manufactures. At any rate, most Yellowstone tourists would have confronted industry as consumers, if not owners of the means of production, and the vistas they met in the park's thermal basins mirrored, in this sense, the landscape of the national economy. The development of Yellowstone, its roads and services, and certainly its promotional literature makes it clear that the nation's first national park truly exerted itself "for the benefit and enjoyment of the people," and that tourists were schooled in the aesthetic consumption of nature's active labor. If Yellowstone itself seemed unable to recover from the curse, if the landscape seemed marred by the endless regeneration and expenditure of force, the nation's touring classes were fast learning how to play as the

consumers of goods produced by others, and the spirit of holiday insured the space between labor and leisure that existed in the culture at large.

If Americans steadily sought out the marketplace of leisure as exemplary citizens of a class who, as Thorstein Veblen has argued, understood "[a]bstention from labor [as] the conventional evidence of wealth and... the conventional mark of social standing," it could be argued as well that they saw their recreation as a defining moment in the narrative of progress, and the result of a national genius which had harnessed the continent's resources to an unprecedented degree.[131] The type of optimism which had been exhibited at the Columbian Exposition in Chicago suggested that the nation had, since its European discovery, commanded the very substance of history, and had carried itself to the far reaches of human potential. Chauncey Depew, delivering a dedication oration for the Exposition buildings, told his audience:

> From the first century to the fifteenth counts for little in the history of progress, but in the period between the fifteenth and the twentieth is crowded the romance and reality of human development. Life has been prolonged and its enjoyment intensified. The powers of the air and the water, the resistless forces of the elements, which in the time of the discoverer were the visible terrors of the wrath of God, have been subdued to the service of man. Art and luxuries, which could be possessed and enjoyed only by the rich and noble,... domestic comforts and surroundings beyond the reach of lords and bishops, now adorn and illumine the homes of our citizens. Serfs are sovereigns and the people are kings. The trophies and splendors of their reign are Commonwealths, rich in every attribute of great States and united in a Republic whose power and prosperity and liberty and enlightenment are the wonder and admiration of the world.[132]

The aesthetic subjugation of the elements in Yellowstone is of a piece with this vision, and to a great extent, the iconology of the park was capable of articulating nearly every aspect of the redemptive scheme as Depew set it forth before the multitude at the Exposition. The great hostelries, the sovereign tourist in the fullness of life, and blessed with the

gifts of the Republic, and not least, the "wonder and admiration of the world," all implied that Yellowstone had yielded its own darker history to the affairs of American progress. Here, broad vistas generated, as they had elsewhere in the Western landscape, the "romance" of human development as a project of redemption, and in the movement from nature's endless labor to the endless day of American leisure lay the transition from profane to sacred, from the wilderness of sin to the culture of grace, and from the Faithful Geyser to the throngs of the geyser faithful, in whose serene, holiday gaze rested the idyll of the Republic.

The majority of Americans in the late nineteenth and early twentieth centuries would have found no real contradiction in allowing the terms of the Christian redemptive scheme to speak so strongly to the ideology of the nation's material progress; Bishop Warren may have discerned more than most with regard to the divine in nature's economy, yet few would have quibbled in principle with the assertion that "[o]n all that God has made he has stamped the great principle of progress, refinement, development."[133] There existed, however, a history of voiced concerns that the narrative of material progress might, in the very magnitude of its realization, supplant the Christian scheme of redemption on its own terms, in a secular rhetoric of paradise. These concerns were certainly in evidence at the Columbian Exposition, where a protracted debate over the opening of the exhibits on Sundays reflected a religious concern that secular achievements might rival even the millennial landscape. Reid Badger notes that "[b]eginning almost as soon as the world's fair opened... the term "new Jerusalem" began to be used to describe the visitor's first impressions," one teacher from Illinois arguing that the "magic splendor of that sight can never be excelled on earth. Will

the New Jerusalem be fairer?"[134] The World's Congress at the Exposition featured a Department of Religion representing "the united forces of religion... rallied against the disintegrating material forces of the age," and director Reverend John Barrows hoped for a "general unanimity of applause that religion shall in some conspicuous way, in this age of materialistic pride, assert its kingship over human life."[135]

If the comparisons of the Exposition's White City with the New Jerusalem seemed a dangerous sort of hubris for the nation to engage in as it examined its achievements, the appearance of a broad culture of leisure making paradisiacal claims upon the nature of experience itself was equally disconcerting. Characterizations of pleasure-seeking as dissipating, as catering to base physicality, or as overtly sinful were hardly new; yet the national sanction of more wholesome leisure activities opened onto a novel field which not only proffered the mark of status for an elect class of citizens, but which suggested that a significant portion of the population might dispense with the conditions of the human struggle and enter into the pursuit of a perfectible human nature. American leisure, while it maintained a respectable distance from openly productive pursuits, nevertheless attached to itself a rhetoric of self-improvement which shaped the entire context of recreational activities. No one was simply free, in other words, to waste the open time with which the Republic had blessed its middle classes; leisure was not a value neutral arena but rather the incipient realm of knowledge, of virtue, and of the body waging war against the effects of time.

In this fashion, the conception of a national park as a source of both "benefit" and "enjoyment" to the American citizenry was of a piece with the broad cultural values surrounding recreation and its role in the

maintenance of the Republic. In Yellowstone, a clear educational and interpretive mission developed slowly, though "cone talks" were given by soldiers of the "Twenty-second Infantry who were stationed in the Upper Geyser Basin in 1888, and educational work of a sort was carried on throughout the stagecoach era by porters of the Yellowstone Park Association, and the part-time guides of the Wylie Camping Company."[136] It seems certain, however, that Yellowstone tourists regarded the Park as an opportunity for self-improvement; by the early 1900's visitors saw interpretive activities as an essential feature of the Park, though initially the position of the Park Service had been one of an "official fear lest suggestion of lessons and study would keep people away from the Parks."[137]

The ideology of self-improvement had hardly been absent from the vision of an American millennium as it continued to be articulated by religious scholars.[138] However, the culture of leisure which advocated social grace, refinement, healthy bodies, excellence in choice pursuits, and a well-defined public arena in which to display these virtues, was not entirely immune from the charges that a secular, more classical form of republicanism commanded the nation's interests in its recreation. The classicist concern with a republican state immune from the effects of time, its worldly pursuit of the universal in the idea of the Good, and its essentially unchristian model of the virtuous citizen had indeed been at the core of what Mark Hanley has described as a Protestant fear that the nation had "assigned redemptive status to republican ideology and American institutions in a secularized eschatology."[139] To the extent that respectable recreation and opportunities for self-improvement defined the arena of leisure in America, the broad range of activities with which the

nation engaged its free-time shaped a complex set of values and goals for such a redemptive vision to set its sights upon. The pursuit of leisure was essentially the pursuit of human potential as a core feature of republican values and purpose.

Of course, recreation maintained strong ties to Christian values, and clearly outdoor activities and sightseeing in parks such as Yellowstone continued to elicit responses in the form of openly religious experiences. Yet Christian and secular holiday often met one another on uncertain terms in the Park, and visitors occasionally found themselves juggling Sabbath consciences with the desire to see the sights. Margaret Cruikshank justified the matter by recalling "the Scotch lassie's saying, 'Sunday does not come aboon the pass" and argued that "here we were surrounded by wonders that we must see then or never," her stay in the Park being almost at an end.[140] Questioning the habits of others was an issue that could be more succinctly stated, however, and Reverend Stanley , who sought at the least to observe Sunday services, stated that at Mammoth he had "preached on Sunday to a clever and very attentive audience," though he lamented that "pleasure-resorts are not the most favorable places to make religious impressions."[141]

Others seemed perplexed that the sights of Yellowstone could elicit anything but a strong religious response, and suggested that tourists ought to view their experiences more as pilgrims than as pleasure-seekers. Carrie Strahorn, for example, quipped that her group left Hell's Half Acre "[f]orming a hundred good resolutions for the rest of our lives," yet no one, certainly, took the tour of the Park as seriously as did Alma White, whose *With God in the Yellowstone* is a remarkable attempt to reassert the Christian redemptive narrative upon a landscape shadowed

by the complaisance of the republican idyll.[142] Originally published in 1920, White's account of her Yellowstone trip betrays the urgency of her own millennial outlook, and she laid claim to the Park's true purpose by stating that it was "necessary at this period of the world's history to have an object lesson like that of the Yellowstone National Park to convince people of the infallibility of God's word." She argued further that "[t]his is the time of fulfilment of prophecy concerning the last days," and the classic harangue of the Jeremiad echoed in her warning that tourists in the basins "learn to magnify His name while suspended from the brittle thread of life over a yawning chasm of burning lava which is threatening to engulf them."[143] Though often inspired, and at moments nearly overcome by elements of the Yellowstone landscape in which the "voice of God in Nature is heard in the smallest whisper, and again in tones of thunder," White clearly understood the Park to be contested ground in the literal Christian drama.[144] Relentless in its depiction of leisure as a culture of complaisance and false security, *With God in the Yellowstone* truly inscribes the topography of the Park with the moral landscape of the nation as she saw it, a "romantic and mysterious region, which in some places seems to be the ante-chamber of heaven and in other places the very mouth of perdition."[145]

White had few questions about the tradition of infernal nomenclature in Yellowstone, and she argued "[t]hat the suggestion of these things should be a mere freak or fancy of the mind is out of the question."[146] And yet the rhetoric of the profane so regularly transgresses the boundaries of the Park that her gaze seems to wander freely over a fallen landscape in which worldliness itself lay under the shadow of the curse. Yellowstone was, to be sure, a vivid templar of Satan's present and

future domain; however, any exegesis of the Park landscape would lead the pilgrim-tourist, White intimated, to the understanding that the narrative of Revelations was to be acted out on the global stage. "Ours is the sphere for which he is contending," she noted, "and where he has so long deceived the nations. Where else could the Angel lock him up but inside the earth?"[147] Clearly, sides were to be taken, and White hinted that if the Yellowstone experience was a turning point in the individual drama of redemption, salvation hung in the balance between faith in the things of this world, and those of the next. "Those who are so fortunate," she argued, "as to see the wonders of the Yellowstone will either soften or harden in their souls, as the result of coming in touch with that which is so closely allied to the supernatural." "God forbid that they should harden," she stated, though the pursuit of leisure as an end unto itself was a distraction in the basins, where, "should there be those who pass by... with cold indifference and a lack of response to what is seen there, it is proof that the world has already played havoc with their finer sensibilities."[148]

The testimony of countless journals and accounts in the history of Yellowstone suggests that "cold indifference and lack of response" was a rare reaction indeed to the spectacle of the Park. Yet White clearly hoped that the Yellowstone experience might become the locus of radical conversions, and of a recommitment to the ongoing work of national, and global redemption. For her own part, after witnessing the sights at the falls, she stated that she "felt renewed inspiration to press the battle against sin and unrighteousness to the gates of heaven and to the depths below," which may account for the zeal with which she tried to club a snake in the Upper Geyser Basin, a practice she engaged in "when it was

possible to do so, thus helping to remove the curse."[149] However, while White shared the conviction that Providence had concealed Yellowstone "from the eyes of civilized man" in order to unveil it at the critical moment in the sacred drama, she felt concern that the American citizenry remained complaisant with regard to the ultimate significance of the Park, and went so far as to suggest that "multitudes from foreign shores [would] swarm like bees within its boundaries and reap the benefits" of what had been intended as "God's free gift to America."[150] The Republic, it seemed, might squander its status as the nation destined to lead the world out of profane, and into sacred history, if only because Americans refused to situate the national landscape in the literal context of the Christian eschatology.

If White had hoped that mass conversions of American tourists in the geyser basins would gird an elect national citizenry for its role in a holy war, she understood, as had Reverend Stanley, that pleasure-resorts and the holiday atmosphere of leisure society were less than conducive to religious impressions. The national exuberance with regard to secular progress, she suggested, threatened to turn the thoughts of American citizens from the ongoing labor of redemption, and she asserted that "[o]ptimism and presumption everywhere characterize the multitude" who had placed their faith too squarely in the Republic. White hinted further that the pursuit of secular comforts resulted a "downward tendency and a proneness to place the affections on material things, to worship the creature rather than the Creator."[151] That such a citizen, epitiomized in the tourist on holiday in the basins, was a poor solider in the war for salvation was clear; White suspected that she herself had nearly fallen

under the spell at the Falls, where "[s]surrounded by nature in all its primitive beauty and grandeur," she stated:

> I forgot my burdens, and seemed to have a foretaste of what it will be when the cross is laid down and the crown is won. But to be an overcomer, I knew there must be no shrinking from duty until the last battle is fought.[152]

Such a statement captures the essential conflict at the heart of the redemptive narratives which, in White's opinion, were in open contest for the Yellowstone landscape. Tourists erred gravely, she suggested, when they placed the sublime and idyllic scenery of the Park within the context of worldliness, or of secular history itself. For insofar as Yellowstone was beautiful, it was not of this world, but the next, and insofar as it was of this world, it "looked as if it might be connected with the place where the king of darkness dwells and his organized forces operate."[153] The pursuit of leisure, she argued, was no arena in which to perfect the human spirit, and the Millennial vacation, which White saw prophesied in the Park laws prohibiting hunting, and in the percieved peaceful coexistence of the wildlife, was no enduring gift of the Republic, but rather an extended moment in the Christian drama, and one which most certainly had a beginning and an end.[154] The inescapable condition of worldliness, she suggested, was writ large in the endless strife of the geyser basins, and the pilgrim tourist did well to cast off the spirit of holiday in the nation's playground. "Many pleasure seekers, she stated:

> Are now thronging to Yellowstone National Park, and in ten thousand demonstrations are coming face to face with the realities of God's word. I saw some standing near the yawning craters of the geysers apparently under deep conviction, and no doubt silently resolving to live different lives, while others frankly stated that the depths of their beings had been stirred and that it was no time to trifle with the soul's eternal interests.[155]

White's Yellowstone jeremiad, then, sounds a late note in the Protestant quarrel with a Republican narrative which, in the view of many divines, had a place in the Christian schema only insofar as it furthered the project of redemption, and only insofar as it was prepared to admit of its time-bound nature and its limitations in alleviating the struggles of the human condition. If the culture of leisure which flourished in the Park encouraged tourists to treat the landscape as the setting for a Far West pastoral, *With God in the Yellowstone* sought to remind them that nature itself labored under the curse in the geyser basins. Yellowstone was exemplary in its sublime scenery, its geology, and even in its celebration of the achievements of civilization as it inscribed upon the landscape the structures of a national playground; yet so too was it singular in that here natural force seemed to shape the face of the earth within a temporal context that confronted the *durée* of human experience itself, and drew the visitor squarely into the heart of natural processes and patterns of regeneration which formed clear limits to the measure of life and the institutions which sheltered it. From this perspective, the Republican narrative offered no guarantee that the nation might, on its own terms, and as a result of its own conquests of the natural world, realize its idyll. Human initiative was, after all, no substitute for the machinations of Providence, and the repose of leisure no real indice of eternity as it was proffered in the Christian eschatology.

Clearly, the onus of representation lies not with the Park itself, but rather, with that "Yellowstone" which Charles Wilkinson suggests was "created out of a human consciousness, at once concrete and profoundly abstract, constructed of geology, history, economics, biology, and spirituality."[156] It goes without saying that such was the complex of

values and attitudes which had shaped the history of expansion and conquest throughout the western landscape, though it seems less certain that a consistent, and progressive ideology yielded a consensus as to the role each of these forces was to play in the nation's development. Certainly the historical and the spiritual consciousness, but so too the biological, geological, and the economic, had each a redemptive context through which to actively, and realistically shape policy and values with regard to the Far West. And while the vast majority of Yellowstone tourists who recorded their experiences did so without viewing themselves as serious commentators or as ideologues, there is little doubt that they understood the Park to be an institution which was of a piece with the national culture, its character, and to a great extent, that sense of destiny and "errand" which Sacvan Bercovitch has argued to be at the core of the American historical consciousness.[157]

To be sure, Yellowstone was handed a role in the national mission through the very language of the act creating the Park as a permanent pleasuring ground, a gesture which itself betrays a confident, and progressive sense of what America's future entailed. In hindsight, the Organic Act was a piece of prophecy for a culture whose citizens were increasingly free to pursue their leisure, and yet the landscape itself must have seemed a promise all along. There is a marker at the junction of the Gibbon and Firehole rivers which consecrates the foresight and patriotism of the Washburn-Doane party as its members wrestled with the implications of what they had seen in the Park; whether the story itself is apocryphal is not the issue so much as the fact that retellings of it found their way into journals and diaries often enough to suggest that American tourists sought in their experiences an intimacy with a national narrative

which, in the case of Yellowstone, had seemed so certain, so progressive, and ultimately, so redemptive in what it offered as a "pleasuring-ground for the benefit and enjoyment of the people."

For Nathaniel Langford, the Yellowstone narrative reasserted the nation's place in the Christian scheme, and its discovery by Angle-European civilization made old things new. "It is a new phase in the natural world," he stated, "a fresh exhibition of the handiwork of the Great Architect."[158] The years following the expeditions of the 1870's would bring huge changes and development to the Park that would introduce the human architect to the landscape, and yet this too was a process of rejuvenation for both wilderness and citizen. Joe Chapple understood the experience at the Upper Basin to be both Alpha and Omega of "the great play that began with the Morning Stars, and passed through the other Lights and Shadows of a Day Eternal." "Here," he suggested:

> before this great palace of rough-hewn logs the christening party is gathered to see the new born day baptized in the waters of a fountain Eternal. To some it is the first, and to others the last act witnessed in the drama of the Yellowstone.[159]

As it blessed the new day, then, Old Faithful blessed the nation, and the vistas which met the tourist in Yellowstone became expansive indeed. Chapple argued that when "one is actually in the Park the geography of the United States looks different. It seems to furnish a bird's-eye-view point that brings the boundary lines of state and nation within quick survey."[160] Chapple, as were countless others no doubt, was pleased with what he saw in his national park survey; the wilderness of the Far West was the scene in which the redemptive drama had come full circle, and he noted that "[i]t is at this time that there comes to us the

theme of Milton--- his visions of Paradise Lost and Paradise Regained."[161] No doubt too he felt that the Yellowstone narrative embraced both secular and sacred history in its idyllic past and future, so long as the Republic sought its Arcadia within the temporal limits of the Christian Millennium. For the true Christian nationalist, the profane landscape of Yellowstone served as a chastening influence upon Republican aspirations too worldly in their designs, and too content in those things that progress could offer the human spirit. For divines like Alma White, the Park remained at best a "piece of Eden preserved upon a thin crust,"[162] and at its worst, in the great geyser basins, an intimation that the human condition itself remained bound to a natural economy that seemed marred by the laws of expenditure and impermanence.

Yet if the Protestant quarrel with the Republic had lodged its complaint in the fear that the nation's fascination with progress paved the way for an assault on the Christian redemptive schema, there remained a field of contention that lay at the heart of a Republic whose economy had ultimately to reject the teleology which had guided the national vision. Christopher Lasch has argued that the confusion of the "modern conception of progress" with the Christian view of history has been misplaced in that the former understood temporal processes "not [within] the promise of a secular utopia that would bring history to a happy ending but the promise of steady improvement with no foreseeable ending at all.... The modern conception of history is utopian only in its assumption that modern history has no foreseeable conclusion."[163] It may be that the secular and the Christian nation parted ways at this juncture of historical thought, though too, it may be as well that here the American citizen

sensed the restlessness of the economic forces shaping the texture and energy of lived experience.

For such a mind, the end of the Yellowstone holiday may have sounded a pensive note in the Republican idyll, and if the Park's profane and laboring landscape suggested to some the condition of worldliness, others bade note that the thermal regions retained a vision that a restless and expansive nation appeared to have foregone. Harry J. Norton, who authored the first real guide-book to the park in 1873, stood at the edge of "one of those calm, lovely prismatic springs" that share the Upper Geyser basin with the more active features, and spoke the equivocal thoughts of the citizen tourist:

> Standing on its margin, one becomes enchanted with this delightful realm, and calls to mind all that he has ever heard or read about the airy phantoms and ogling water sprites of mythological tales. It seemed as if our eyes would never cease feasting upon its unearthly beauty; there was an intense longing to know what mysterious treasures lie hidden deep down in its tranquil bosom; and as we regretfully, unsatisfiedly retired to rest, we wondered who would care if we explored its deepest recesses--- to return to earth never again![164]

Norton had no room in the Old Faithful Inn to retire to, and yet his dreams carried him deep into "Circe's Boudoir," and its "mossy beds softer than eider down."[165] Yellowstone's most compelling idyll, it seems, remained an otherworldly construct. The Yellowstone tourist, on the other hand, would continue to find ways to make the Park both blessing and curse within the broad vistas of the national landscape. The rhetoric of ambivalence has indeed shifted its ground to a modern series of complaints about crowds, costs, transportation, services, authority, and certainly, policy. And yet these issues had always been present in some form or another, and they had always been viewed as problems that cut squarely into the spirit of holiday Americans sought with their free time. The question then, is one of the manner in which the wilderness pastoral found itself displaced within a restless narrative more committed to process than to eschatology. The Yellowstone landscape itself, it could be argued, has been made to commit to a concept of the wilderness idyll

constructed as a pause, or breach in a drama of endless expansion. The American holiday maintained its ties to the vision of leisure as an experience "out of time," as it were, and yet the tourist hoping for rejuvenation sought leisure as a means to return to, not escape from, the processes of history. For the consciousness both drawn to and distracted by the wilderness pastoral, the blessings of eternity remained disconcerting against the forces driving process In Yellowstone.

---

[1] The Line comes from William Tod Helmuth's poem "The Yellowstone National Park and How It Named," printed in Alma White, *With God in the Yellowstone.* (Zarepath, NJ: Pillar of Fire, 1933), pp. 134-37, quoted p. 136.

[2] Nathaniel Pitt Langford, *The Discovery of Yellowstone National Park.* (Lincoln, NE: University of Nebraska Press, 1972), p. 28. It should be noted that this published account, while written in the form of a diary, is "essentially a reminiscent account drawn from diary and other sources." See Aubrey L. Haines' "Forward," pp. vii-xxi, cited p. vii.

[3] Charles Dudley Warner, "Editor's Study." *Harper's*, January 1897, pp. 320-25. Reprinted in *Old Yellowstone Days*. ed. Paul Schullery (Boulder, CO: Colorado Associated University Press, 1979), pp. 159-71, quoted p. 161. The Schullery text is hereafter cited as *OYD.*

[4] *ibid.*, p. 47.

[5] Langford's "mingled dread and wonder" evokes the "aesthetics of the sublime," articulated by Burke, and widely utilized in literature of the American West. A good overall discussion can be found in Kris Lackey, "Eighteenth-Century Aesthetic Theory and the Nineteenth-Century Traveler in Trans-Allegheny America: F. Trollope, Dickens, Irving and Parkman, in *American Studies* 32 (1) (Spring, 1991), pp. 33-48.

[6] Hiram Martin Chittendon, *The Yellowstone National Park.* (Cincinnati, OH: The Robert Clarke Company, 1909), p. 47. Chittendon was also an officer of the Corps of Engineers who is credited with much of the pioneering construction of Yellowstone National Park's road system in the late Nineteenth, and early Twentieth Centuries.

[7] *ibid.*, pp. 50-51, and p. 49, respectively. The "mountain of transparent glass" is the Obsidian Cliff, located between Mammoth Hot Springs and the Norris Geyser Basin.

[8] *ibid.*, p. 47. Ferdinand Hayden, however, who was to lead a Geological Survey party into the area in 1871, noted with pleasure Bridger's storytelling. Hayden had been a member of the 1860 Raynolds expedition which had been guided by Bridger, and though the expeditin failed to make its way into the headwaters of the Yellowstone, Hayden stated that "Bridger, the best mountain man the west has ever produced, was quite familiar with the country, and frequently entertained us with marvelous descriptions of the wonders of that great Basin." See "The Hot Springs and Geysers of the Yellowstone

and Firehole Rivers." *The American Journal of Science and Arts* Third Series, vol. III no. 13 (February, 1872), pp. 105-15, quoted p. 105.

[9] Aubrey L. Haines, *The Yellowstone Story: A History of Our First National Park.* 2 vols. (Yellowstone National Park: Yellowstone Library and Museum Association, 1977), vol. 1, p. 137. Haines' two volume history of Yellowstone is the best and most comprehensive treatment of the park's history and development, and I am indebted to it throughout this chapter for both general and specific information. Haines notes that Langford was to deliver a series of twenty lectures as part of this commitment, though the series was canceled for reasons not entirely clear. See Haines vol. 1, p. 140.

[10] See Haines, *The Yellowstone Story,* vol. 1, Chapter 6, "The New Creation," pp. 156-73 for a brief overview of issues relating to the Northern Pacific Railroad's interest in the park.

[11] Quoted in Haines, *The Yellostone Story,* vol. 1, p. 100. An account of the Cook-Folsom expedition, with further information regarding the press's reluctance to publish early accounts of Yellowstone, is found in vol. 1, pp. 91-101.

[12] Haines, *The Yellowstone Story,* vol. 1, pp. 134-38. Also of note here is Truman C. Everts' "Thirty-seven Days of Peril," *Scribner's Monthly* 3 (November 1871), pp. 1-17. Everts, another member of the Washburn- Doane party, was separated from the main group, and wandered lost in the park until rescued by George Pritchett and John Baronett. See Haines, *The Yellowstone Story,* vol. 1, 125-34.

[13]William H. & William N. Goetzmann, *The West of the Imagination.* (New York: W.W. Norton & Company, 1986), p. 171. See "Artist and Photographer in Wonderland," pp. 170-82 for an account of Moran's and Jackson's experiences with Hayden's expedition, as well as public response to and marketing of Yellowstone images by both artists.

[14] Chittendon, *The Yellowston National Park*, p. 48.

[15] Barbara Novak, *Nature and Culture: American Landscape and Painting, 1825-1875.* (New York: Oxford University Press, 1980), p. 7.

[16] Ferdinand V. Hayden, "Hot Springs," p. 108. And Margaret Andrews Cruikshank, "Mortal Language Seems Entirely Inadequate: A Lady's Trip to Yellowstone in 1883," ed. Lee H. Whittlesey, manuscript courtesy of Lee Whittlesey, Yellowstone National Park Library. Entry for August 24, 1883. A shorter version of this manuscript can be found in *Montana Magazine of Western History* 39: 2-15, Winter, 1989.

[17] Anne Farrar Hyde, *An American Vision: Far Western Landscape and National Culture, 1820-1920.* (New York: New York University Press, 1990), p. 7.

[18] Carrie Adell Strahorn, *Fifteen Thousand Miles By Stage.* (New York: G.P. Putnam's Sons, 1911), p. 272. Strahorn's trip was made in the fall of 1880.

[19] Novak, *Nature and Culture*, p. 7.

[20] Quoted in Perry Miller, *Errand into the Wilderness.* (Cambridge, MA: Harvard University Press, 1956), p. 210.

[21] Warner, *OYD*, p. 162.

[22]*ibid.*, p. 161.

[23] Organic Act of 1872, quoted in Haines, *The Yellowstone Story*, vol. 2, p. 471.

[24] Langford recounts the story as follows: "Last night, and also this morning in camp, the entire party had a rather unusual discussion. The proposition was made by some member that we utilize the result of our exploration by taking up quarter sections of land at the most prominent points of interest, and a general discussion followed. One member of our party suggested that if there could be secured by pre-emption a good title to two or three good quarter sections of land opposite the lower fall of the Yellowstone and extending down the river along the canyon, they would eventually become a source of great profit to the owners. Another member of the party thought that it would be more desirable to take up a quarter section of land at the Upper Geyser Basin, for the reason that that locality could be more easily reached by tourists and pleasure seekers. A third suggestion was that each member of the party pre-empt a claim, and in order that no one should have an advantage over the others, the whole should be thrown into a common pool for the benefit of the entire party.

Mr. Hedges then said that he did not approve of any of these plans--- that there ought to be no private ownership of any portion of that region, but that the whole of it ought to be set apart as a great National Park, and that each one of us ought to make an effort to have this accomplished. His suggestion met with an instantaneous and favorable response from all --- except one--- of the members of our party, and each hour since the matter was first broached, our enthusiasm has increased." Langford, *Discovery of Yellowstone*, pp. 117-118. See Haines, *The Yellowstone Story*, vol. 1, pp. 129-131 for an account of Langford's story and Haines' suspicions regarding its truth. See, too, his introduction to Langford, *Discovery of Yellowstone*, pp. vii-xxi.

[25] Rudyard Kipling, *From Sea to Sea; Letters of Travel.* 2 vols. (New York: Charles Scribner's Sons, 1906), vol. 2: 136-89. Reprinted in *OYD*, pp. 85-114, quoted p. 98.

[26] Mrs. N.E. Corthell and Mrs. John A. Hill, *A Family Trek to the Yellowstone, and Twenty-four Years After.* (Laramie, WY: Laramie Printing Co., 1923), pp. 36-37. Excerpt courtesy of Yellowstone National Park Library.

[27] Dr. Wayland Hoyt, in *Elbert Hubbard's Scrapbook.* (New York: Wm. H. Wise & Co., 1923), pp. 28-29. Excerpt courtesy of Yellowstone Natrional Park Library.

[28] Mary Bradshaw Richards, *Camping Out in the Yellowstone,* ed. William W. Slaughter. (Salt Lake City: University of Utah Press, 1994), p. 88.

[29] See Roderick Nash, *Wilderness and the American Mind.* (New Haven: Yale University Press, 1967), esp. "The Romantic Wilderness," pp. 44-66. For a discussion of the Pastoral as a mediating, or "middle landscape" between unspoiled nature and high civilization, See Leo Marx, *The Machine in the Garden: Technology and the Pastoral Ideal in America.* (New York: Oxford University Press, 1964.), esp. "The Garden," pp. 73-144.

[30] Edwin J. Stanley, *Rambles in Wonderland.* (New York: D. Appleton & Company, 1880), p. 10. Yellowstone National Park Library. The biblical reference is to Isaiah 35:1.

[31] John F. Sears, *Sacred Places: American Tourist Attractions in the Nineteenth Century.* (New York: Oxford University Press, 1989), p. 162.

[32] John Muir, "The Yellowstone National Park." *Atlantic Monthly*, April 1898, pp. 509-22. Reprinted *OYD*, pp. 41-63, quoted p. 42.

[33] Stanley, *Rambles in Wonderland*, p. 150.

[34] *ibid.*, p. 150. The cayuse forms a central topic of discussion in many early accounts which, while praising this horse's abilities, in general amount to something of a character assassination for which Stanley said "I will not apologize too much, for more than once have I been the unfortunate and unhappy victim of these creatures while exhibiting their

striking propensities in a very forcible manner.... Pray to be delivered from such misfortunes when you come West!" , pp. 126-27.

[35] Glen Uminowicz, "Recreation in a Christian America: Ocean Grove and Asbury Park, New Jersey, 1869-1914," in Kathryn Grover, ed., *Hard at Play: Leisure in America, 1840-1940.* (Amherst, MA: University of Massachusetts Press, 1992), pp. 8-38, quoted p. 23.

[36] Muir, *OYD*, p. 62.

[37] *ibid.*, pp. 53-54.

[38] Sears, *Sacred Places*, p. 174.

[39] Muir, *OYD*, p. 60, p. 62.

[40] Kipling, *OYD*, p. 109. Yellowstone Park Archivist Lee Whittlesey is among a number of scholars who reject the idea that Native Americans avoided the Yellowstone area out of fear. Noting a distinction between fear and reverence, he suggests that this folklore was introduced essentially as an interpretive motif evoking a safe image of the Park by Philetus W. Norris, Superintendent of the Park from 1877-1882 (personal interview). Alston Chase argues more acerbically that the myth also served to deny the human past of Yellowstone, an effort to dispel agency in what was essentially viewed as a pristine, natural area. See Alston Chase, *Playing God in Yellowstone.* (New York: Harcourt Brace & Company, 1987), esp. "Growing Apples in Eden," pp. 92-115.

[41] *ibid.*, p. 109.

[42] *ibid.*, p. 89.

[43] *ibid.*, p. 89.

[44] Nash, *Wilderness and the American Mind*, p. 68. See Novak, *Nature and Culture*, "The Geological Timetable: Rocks," pp. 47-77.

[45] Any study of the concept of the New World relation to sacred history would do well to begin with Perry Miller's *Errand into the Wilderness*, especially the title essay, pp. 1-15. Sacvan Bercovitch's *The American Jeremiad.* (Madison, WI: University of Wisconsin Press, 1978) argues (most cogently as a response to Miller's essay) the tenacity of the American concept of mission, and its appropriation of the typology of sacred history as the means of interpreting and shaping its temporal affairs.

[46] Nicholas Mohr, *Excursion Through America*, ed. Ray Allen Billington, tr. La Vern J. Ripley and Klaus Lanzinger. (Chicago: R.R. Donnelley & Sons Company, 1973), pp. 269-70. Excerpt courtesy of Yellowstone National Park Library.

[47] Sidford Hamp, "Exploring the Yellowstone with Hayden: Diary of Sidford Hamp," ed. Herbert Oliver Brayer, *Annals of Wyoming* 14, no. 4 (October, 1942), pp. 253-298, quoted p. 269. Yellowstone National Park Library. Hamp was the Nephew of William Blackmore, English lawyer and financier, who had intended as an acquaintance of Hayden to make the journey as well. His wife Mary, however, became ill and died in Bozeman; Blackmore turned back, leaving his nephew to travel on alone with the group. See Brayer's footnote no. 32, p. 279.

[48] Theodore Roosevelt, "Wilderness Reserves," *Forest and Irrigation* 10 (June 1904), pp. 250-59; (July, 1904), pp. 300-09. Reprinted in *OYD*, pp. 185-205, quoted pp. 186-87. Roosevelt's account of Yellowstone is somewhat unique in that he is, for the most part, interested in the area's game population; the account is, essentially, a series of mock

stalkings of various of the Park's elk, antelope, and bighorn sheep herds.

[49] Kipling, in Schullery, *OYD*, p. 88. For a discussion of tourism in Yellowstone in the late nineteenth, and early twentieth Centuries, see Haines, *The Yellowstone Story*, "On the Grand Tour," vol. II, pp. 100-59. See also Richard Bartlett, *Yellowstone: A Wilderness Besieged.* (Tucson, AZ: University of Arizona Press, 1985), esp. "The Railroad Age,: 1883-1915," pp. 43-72.

[50] *ibid.*, p. 88.

[51] *ibid.*, p. 89.

[52] *ibid.*, p. 89.

[53] See Bercovitch, *American Jeremiad*, esp. "Ritual of Consensus," pp. 132-175.

[54] Rush Welter, *Mind of America, 1820-1860* (New York, 1975), cited in Bercovitch, *American Jeremiad*, p. 145.

[55] Kipling, *OYD*, p. 89.

[56] *ibid.*, p. 89. Kipling's "excellent Rayment" is Raymond Excursions, an arm of the Raymond and Whitcomb Travel Agency of Boston which began arranging tours through the Park in the 1890's.

[57] Frederick Jackson Turner, "The Significance of the Frontier in American History," in *The Frontier in American History*. (Huntington, NY: Robert E. Krieger Publishing Company, 1976), pp. 1-38, quoted, p. 38. A good deal of recent scholarship has taken up a critique Turner's thesis, noting as does Patricia Nelson Limerick that "[t]here is simply no definition of 'the closing of the frontier' that is anything but arbitrary and riddled with exceptions and qualifications." Turner's thesis, she points out, rests too fundamentally upon a spatial/agrarian model of westward expansion that denies the continuity of other characteristic forms of development in the West, such as mining, railroads, urbanization, commerce, finance, etc. It is important to note, however, that Nelson Limerick's critique centers most squarely upon the views of historians themselves, and that Americans have in actuality had little trouble in viewing the West as an expanding realm of opportunity and continued development. See Patricia Nelson-Limerick, *The Legacy of Conquest: The Unbroken Past of the American West.* (New York: W.W. Norton & Company, 1987), quoted p. 23.

[58] Kipling, *OYD*, p. 89.

[59] Walter Raymond and I.A. Whitcomb, *Five Grand Trips to the Yellowstone National Park.* (Boston, MA: 1890), p. 4.

[60] Mohr, *Excursion*, pp. 256-59.

[61] Washington Irving, *A Tour on the Prairies*, ed. John Francis McDermott. (Norman, OK: University of Oklahoma Press, 1956), p. 4, and p. 55, respectively.

[62] Kipling, *OYD*, p. 102.

[63] Stanley, *Rambles in Wonderland*, p. 155. Osborne Russell had, as early as 1836, surmised of one steam vent on the northern extremity of Yellowstone Lake that "the steam issued with sufficient force to work an engine of 30 horsepower." *Journal of a Trapper*, ed. Aubrey L. Haines. (Lincoln, NE: University of Nebraska Press, 1955), p. 44.
[64] Mircea Eliade, *The Sacred and the Profane: The Nature of Religion*, tr. Willard R.

Trask. (New York: Harper & Row, 1961), p. 36.

[65] Stanley, *Rambles in Wonderland*, p. 156.

[66] *ibid.*, p. 78.

[67] See, in addition to Bercovitch, *American Jeremiad*, Ernest Lee Tuveson, *Redeemer Nation: The Idea of America's Millennial Role*. (Chicago: University of Chicago Press, 1968). Any study of Republican thought should begin with J.G.A. Pocock, *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. (Princeton: Princeton University Press, 1975), esp. with relation to American thought "The Americanization of Virtue," pp. 506-552. See also Joyce Appleby, *Liberalism and Republicanism in the Historical Imagination*. (Cambridge: Harvard University Press, 1992).

[68] Frederic Jameson, *The Political Unconscious: Narrative as a Socially Symbolic Act.* (Ithaca, NY: Cornell University Press, 1981), p. 19.

[69] Tuveson, *Redeemer Nation*, p 194. Tuveson is responding here to Hollis Read's *The Coming Crisis of the World*, written in 1861. This is certainly not the only set of images available to Americans trying to conceptualize millennial society. Tuveson effectively argues, however, the pervasive "this-worldly character" of true millennial thought, and even more primitivist, "Edenic" models are often marked by appeals to agricultural and technological development yielding a sort of "fertility miracle" of abundance. See pp. 58-73.

[70] Henry White Warren, *Among the Forces.* (New York: Eaton & Mains, 1898), pp. 157-158. Yellowstone National Park Library.

[71] *ibid.*, p. 115, and p. 144, respectively.

[72] *ibid.*, p. 77.

[73] Warner, *OYD*, p. 161. John Burroughs, from *Camping and Tramping with Roosevelt* (Boston: Houghton, Mifflin and Co., 1907), a Yellowstone account reprinted in *OYD,* pp. 209-228, quoted p. 224.

[74] Cruikshank, "Mortal Language," entry for August 24.

[75] Owen Wister, in *Harper's*, March, 1936, reprinted in OYD, pp. 67-83, quoted pp. 72-73.

[76] Hughes, Thomas McKenny and Mrs. Caroline, "Excerpts from the Hughes Diary While Attending An International Geological Congress in America in 1891." Entry for September 6. Yellowstone National Park Library.

[77] Hannah Arendt, *The Human Condition.* (Chicago: The University of Chicago Press, 1958), p. 85.

[78] *ibid.*, p. 81.

[79] Lee Whittlesey, *Yellowstone Place Names*. (Helena, Montana: Montana Historical Society Press, 1988), p. 70. This work is an abbreviated version of Whittlesey's *Wonderland Nomenclature*, which is much more extensive, and should be regarded as the definitive source on the history of nomenclature in Yellowstone.

[80] Warren, *OYD*, p. 10.

[81] Stanley, *OYD*, pp. 107-108.

[82] *ibid.*, p. 108.

[83] Carol Merchant, *The Death of Nature: Women, Ecology, and the Scientific Revolution.* (New York: Harper & Row, 1980), p. 8.

[84] Stanley, *Rambles in Wonderland*, p. 108.

[85] Sears, *Sacred Places*, p. 8. See the "Introduction" for Sears' thoughts on the culture of American tourism and gender. See, for a discussion on the femininization of the American landscape, Annette Kolodny, *The Lay of the Land: Metaphor as Experience and History in American Life and Letters*. (Chapel Hill: University of North Carolina Press, 1975).

[86] Dorothy Brown Pardo, "Dorothy in Wonderland," Personal account of a 1911 trip to Yellowstone, p. 28. Manuscript, Yellowstone National Park Library.

[87] See, for example "The First Fast Food?" in *National Geographic* April, 1996, p. 134.

[88] Margaret Andrews Allen, "A Family Camp in Yellowstone Park," *Outing* 7 (Nov. 1885), pp. 157-59, quoted p. 158.

[89] Joe Mitchell Chapple, *A'Top O' the World: Wonders of the Yellowstone Dreamland.* (Boston: Chapple Publishing Co., 1922), p. 70. Yellowstone National Park Library. For the basis of this story in a factual event, see Haines, *The Yellowstone Story*, vol. 2, p. 17.

[90] Stanley, *Rambles in Wonderland*, p. 107.

[91] See Whittlesey, *Yellowstone Place Names*, p. 70. Whittlesey notes that the pool became dormant sometime between late 1926and early 1927.

[92] Warner, *OYD*, pp. 161-62.

[93] Cruikshank, "Mortal Language," entry for August 26, 1883.

[94] Hamp, "Diary," "The Tail of a Shirt," pp. 280-83.

[95] *ibid.*, p 279-280

[96] See , for example, Chase, *Playing God in Yellowstone*, esp. "The Deep Hole Gap," pp. 262-92.

[97] For an account of the 1872 expedition see Haines, *The Yellowstone Story*, vol. 1, pp. 181-92.

[98] See Lee H. Whittlesey, *Death in Yellowstone:Accidents and Foolhardiness in the First National Park*. (Boulder,CO: Roberts Rinehart Publishers, 1995), for a discussion of human mortality in the park.

[99] Corthell, Mrs. N.E. and Hill, Mrs. John A., *A Family Trek to Yellowstone*, p. 24.

[100] Kipling, *OYD*, p. 98, and p. 102.

[101] Henry Davenport Northrop and Nancy Houston Banks, *The World's Fair as Seen in One Hundred Days.* (Cincinnati, OH: The Standard Publishing Co., 1893), p. 232. The authors agreed with nearly everyone that the Machinery Hall in the Fair's White City

represented the "Marks of the World's Progress" and "Stupendous Mechanical Energy." The Hall was built, they asserted, was "especially constructed for exhibiting machinery in motion," a quality which made it "a live exhibit, differing in this respect from almost every other department." "Within its massive walls," they stated, "was the ceaseless hum of machinery in motion," and they suggested that the popularity of the exhibit lie in the fact that "[t]he fascination of moving machinery is more than an attraction to the skilled artisan merely. It draws the general public...." See "Machinery Hall," pp. 229- 38. For a discussion of the World's Columbian Exposition as a model future society and mechanized utopia, see Allen Trachtenburg, *The Incorporation of America: Culture and Society in the Gilded Age*. (New York: Hill And Wang, 1982), esp. "White City," pp. 208-34. See also Justus D. Doenecke, "Myths, Machines and Markets: The Columbian Exposition of 1893," in *Journal of Popular Culture* (Winter, 1972), pp. 535-49.

[102] "The Power Plant," *Harper's Weekly* XXXVII (May 13, 1893), p. 442.

[103] "Opening of the Great Fair," *Harper's Weekly* XXXVII (May 13, 1893), p. 442. These remarks are attributed to "the blind chaplain of the Senate, Rev. Mr. Milburn," in his dedicatory prayer.

[104] Kipling, *OYD*, pp. 110-11.

[105] Eliade, *Sacred and Profane*, p. 87.

[106] Though we shall have occasion to return to this issue, it will be noted here that the Park's visible wildlife suggested to more than one visitor a vision of the millennial landscape; Joe Mitchell Chapple, for example, who stated that at "Yellowstone, the calendar (*sic*) is forgotten; time is measured only by the moods of Mother Nature that diffuse through the very being of man and give to him that unconscious knowledge that the hours are passing for Time and Eternity," argued as well that "under the spell of a natural, unrestrained goodness, all creatures and even the forests that stretch themselves out under the canopy of the heavens, seem to be imbued with a spirit of goodness and godliness." He noted that "[h]ere it is that man and beast know and love each other; the elk, the bear, the deer, the moose, the buffalo and the jack-rabbit, the woodchuck and the chipmunk--- all make their homes together in peace, here in the playground of men." Chapple, *A'Top O' the World*, pp. 61 and 63.

[107] Olin D. Wheeler, *Wonderland 1904*. (St. Paul: Northern Pacific Railway, 1904), p. 33. Haynes Special Collections, Renne Library, Montana State University.

[108] See Arendt, *The Human Condition*, esp. "Work," pp. 136-174. Arendt argues that *homo faber*, whose capacity for work arises out of the surplus capacity of labor in its reproduction of the life process, erects the world of human artifice as a shield against the cyclicality of natural processes, within which human life ultimately finds its limitations and its fundamental realm of necessity, since life must, in essence, reproduce itself daily.
[109] Farrar Hyde, *An American Vision*, p. 256.

[110] Anonymous, in *Western Architect* magazine, quoted in Farrar Hyde, *An American Vision*, p. 257. Wallace Stevens, "Anecdote of the Jar."

[111] Sanders, Wilbur Edgarton, "Journal of a Trip to Yellowstone, August 19- September 8, 1881." Manuscript ,Yellowstone National Park Library, entry for August 28, at the Upper Basin.

[112] Chapple, *A'Top O' the World*, p. 65.

[113] *ibid.*, p. 65.

[114] *ibid.*, pp. 65-66.

[115] William Tod Helmuth, "The Yellowstone Park and How it Was Named," reprinted in Alma White, *With God in the Yellowstone.* (Zarepath, NJ: Pillar of Fire, 1933), pp. 134-37, quoted p. 136.

[116] Stanley, *Rambles in Wonderland*, p. 66.

[117] Corthell, Mrs. N.E. and Hill, Mrs. John A., *A Family Trek to Yellowstone*, p. 32.

[118] See Whittlesey, *Death In Yellowstone.*

[119] Hughes, Thomas McKenny and Mrs. Caroline, "Excerpts from the Hughes Diary While Attending an International Geological Congress in America in 1891," entry for September 10, 1891.

[119] Hague, letter to R.B. Moore, December 14, 1908, National Archives, U.S. Geological Survey, Record Group 57, Hague Papers, Entry 68. A brief discussion of Hague's career can be found in Bartlett, *A Wilderness Besieged*, pp. 142-43. Haines notes a "Hell-Broth Springs" located in the Sulphur Creek drainage at the base of Mount Washburn, and named by members of the Washburn-Doane Expedition. *The Yellowstone Story*, vol. 1, p. 116.

[121] The idea of Yellowstone as artifice is perhaps the central topic of debate with regard to park policies and values in the administration of a complex ecosystem. Alston Chase's *Playing God in Yellowstone* is the most familiar and thoroughgoing critique of the Park Service's assertion that Yellowstone ought to be treated as a natural, intact ecosystem left to its own processes, while the Park Service itself has wrestled thoughtfully with these issues , most notably, perhaps in its role in the 1991 Vail symposium marking the Park Service's 75th Anniversary. See *National Parks for the 21st Century: The Vail Agenda* (Post Mills, Vermont: Chelsea Green Publishing Company). This debate is not really of concern in a study of the history of tourism in Yellowstone, which is in many facets the history of its accessibility, its system of roads and pathways shaping the park experience, and the construction of spaces fit for human habitation both inside and adjacent to park boundaries. The Park Service has roughly 36,00 square miles of land to administer, and to be sure, the future of development, policies regarding existing structures and features, visitor density concerns, etc. all form part of a valid and *current* debate; yet to argue against the nature of human activity as a vital shaping force on the landscape of Yellowstone is simply untenable. The vast, vast majority of tourists *experience* the park as a construct of various ideologies of leisure, of economic forces, and of contextualized confrontations with the natural world.

[122] The act of recording the day's experiences was itself a social assertion of leisure according to Kipling, who noted that after his fellow tourists had seen the sights in the Upper Geyser Basin, they all "returned to the hotel to put down their impressions in diaries and note-books which they wrote up ostentatiously in the verandahs." In *OYD.*, p. 106.

[123] Chapple, *A'Top O' the World*, p. 68.

[124] Wheeler, *Wonderland 1904*, "The Lignite Coal Fields of North Dakota," pp. 21-28. Quoted, p. 22, p. 21, & p. 28, respectively.

[125] *ibid.*, p. 22.

[126] Reid Badger, *The Great American Fair: The World's Columbian Exposition and*

American Culture. (Chicago: Nelson Hall, 1979), p. 28.

[127] Wister, *OYD*, p. 71.

[128] Rebecca Harding Davis, *Life in the Iron Mills*, in *The Norton Anthology of American Literature*, Shorter Edition. (New York: W.W. Norton & Co., 1989), pp. 1123-49, quoted p. 1128.

[129] Corthell, Mrs. N.E., and Hill, Mrs. John A., *A Family Trek*, p. 32.

[130] Sears, *Sacred Places*, p. 183.

[131] Thorstein Veblen, *The Theory of the Leisure Class: An Economic Study of Institutions.* (New York: The Modern Library, 1934), p. 41.

[132] Quoted in Northrop and Banks, *The World's Fair*, p. 185.

[133] Warren, *OYD*, p. 173.

[134] Badger, *The Great American Fair*, p. 97. Mrs. D.C. Taylor, *Halcyon Days in the Dream City* (Kankakee, Illinois: 1894), P.7, cited in Badger, p. 97.

[135] quoted in Badger, *The Great American Fair*, p. 99.

[136] Haines, *The Yellowstone Story*, vol. 2, p. 307. See pp. 303-309 for a history of interpretive/educational programs in the park. The Wylie Camping Company, which operated "permanent" tent camps in the Park beginning in 1893, "opened the park to people of lesser means," yet provided middle class entertainment and an "effort to supply [its] guests with accurate, interesting information about the Park." See, vol. 2, pp. 136-41.

[137] Quoted in Haines, *The Yellowstone Story*, vol. 2, p. 303.

[138] See Tuveson, *Redeemer Nation*, pp. 59-63.

[139] Mark Y. Hanley, *Beyond a Christian Commonwealth: The Protestant Quarrel with the American Republic, 1830-1860.* (Chapel Hill: University of North Carolina Press, 1994), p. 3.

[140] Cruikshank, "Mortal Language," entry for Sunday, August 26. Lee Whittlesey's footnote #61 states "This Scottish saying means that Sunday rest cannot be enjoyed before the efforts of the week are over."

[141] Stanley, *Rambles in Wonderland*, p. 61.

[142] Strahorn, *Fifteen Thousand Miles*, p. 61.

[143] Alma White, *With God in the Yellowstone*, p.90, and p. 130.

[144] *ibid.*, p. 49.

[145] *ibid.*, p. 41.

[146] *ibid.*, p.121.

[147] *ibid.*, p. 115.

[148] *ibid.*, pp. 128-30.

[149] *ibid.*, p. 48, and p. 112.

[150] *ibid.*, p. 41.

[151] *ibid.*, p. 90, and p. 132.

[152] *ibid.*, p. 53.

[153] *ibid.*, p. 107.

[154] *ibid.*, see p. 52. White quotes from Isaiah 11:5-9.

[155] *ibid.*, p. 91.

[156] Charles F Wilkinson, *The Eagle Bird: Mapping a New West.* (New York: Vintage Books, 1993), p. 170.

[157] See Bercovitch, *American Jeremiad*, "The Puritan Errand Reassessed," pp. 3-30.
[158] Langford, *Discovery of Yellowstone*, p. 122.

[158] Chapple, *A'Top O' the World*, p. 79.

[159] *ibid.*, p. 79.

[160] *ibid.*, p. 81.

[161] *ibid.*, p. 53.

[162] White, *With God in the Yellowstone*, p. 56.

[163] Christopher Lasch, *The True and Only Heaven: Progress and its Critics*. (New York: W.W. Norton & Co., 1991), pp. 47-48.

[164] Harry J. Norton, *Wonder-Land Illustrated*. (Virginia City, MT: Harry J. Norton, 1873), p. 18. Yellowstone National Park Library.

[165] *ibid.*, p. 21.

CHAPTER 4

"IF GOD MADE FISHIN', A FELLER ORTER FISH": GRACE, LEISURE, AND THE CULTURE OF FLY FISHING

> As someone once tried to explain it (with an angler's typical enthusiasm): "God does not deduct from a man's time on earth the hours he spends fishing."
>
> George X. Sand[1]

There is something in this assertion of Sand's which begs the question of the temporal nature of leisure; for if, indeed, hours are spent fishing, if , in other words, the clock continues to tick for the angler casting a line to a fish, from whence does such time get deducted, except from eternity itself? Of course expenditure is meaningless here with regard to time; it is always already *redeemed*, as it were, inasmuch as it draws on the plenitude of eternity. It is, however, fraught with meaning for the individual whose leisure comes to be the indice of an experience spent altogether outside of time, as if the gift of a few hour's eternity on earth were itself the sign and seal of redemption--- Christian grace made fluid in the image of the body beautiful casting tight loops in the middle of a river.

I don't know whether Sand was an angler or not. The "angler's typical enthusiasm" isn't necessarily shared by any individual who fishes, which is probably a good thing. I do wonder, however, what Sand would think today of that stretch of the Yellowstone river known as Buffalo Ford, where, in the middle of the nation's most cherished national park, in the months between June and early September, the culture of American fly fishing convenes in a mass spectacle that can only be described as a yearly "great reawakening" of faith in the value of leisure, a faith which must be great, indeed, in such surroundings. Home to an indigenous

population of Yellowstone cutthroat trout, and temporary residence of a transient population of anglers from all over the world, it's unclear what notions of eternity might be wrought from this stretch of water where time and space are at such a premium, and where, if God is not keeping track of the hours, someone at the Park Service is.[2] The iconology of fly fishing seems undone here, where the physical body is more an assertion toward the "angling Other," than the reflection of a specific posture taken toward the natural world, toward the flux of time. Casting a fly in such an environment is hardly an evocation of eternity; it is, rather, an incessant limning of personal space, a relational stance one individual takes toward another in the midst of what everyone understands to be a finite set of values--- it's a startling claim for the right to expend time, where the places in which to expend it have become so competitive.

This is no idyll for the lost age of the solitary angler; lamentations on the "growing popularity" of the sport, and the loss of privacy on good stretches of water are hardly new, and have often lead to the privatization of resources; the assertion of the right to fish privately has a troubling history of laying claim to the natural world at the expense of others.[3] There are solitary stretches of water to be found in the park, if one so chooses; there are grizzly bears there too. What interests me, on the other hand, with respect to the famous, and populated stretches of the Yellowstone and other great rivers of the west, is the manner in which they construct this equivocal ground of appearance in what is essentially a social activity; seeing others, and being seen, is fundamental to the sport no matter what anyone asserts to the contrary. If fly fishing is not public, not openly visible to others by virtue of a social gathering of anglers, then it is publicized, becomes text--- written about, talked about, remembered.

Charles Cotton's portion of the *Compleat Angler, or, The Contemplative Man's Recreation*, remains profoundly insightful on the social nature of angling, and though no one would expect to hear the language spoken on the river Dove to be overheard on the banks of the Yellowstone, one suspects the tenor of this piece of angling *logos* remains intact wherever its sentiments are spoken. The narrative of Cotton's text takes the form of a Socratic dialogue between a knowledgeable *Piscatore Junior*, and one less learned angler, *Viatore*. The two are waiting out a hatch of mayflies on the river, and the *Piscatore* gives to his friend the understanding that, in case the fishing is slow, he is always willing to talk:

> *Pisc*. I have told you how a flie is to be made, you shall presently see me make one, with which you may peradventure take a trout this morning, notwithstanding the unlikeliness of the day; for it is now nine of the Clock, and fish will begin to rise, if they will rise to day; I will walk along by you, and look on, and after dinner I will proceed in my lecture of Flie-Fishing.

The *Viatore* is willing to express an urge to be fishing; however, he suggests as well that "yet I could sit here all day to hear you; but some of the one, and some of the other will do well." The *Piscatore* has concerns about the fishing, and yet suggests he will find a way to get the *Viatore* one good day on the river within the month, if at all possible. Trading on the fantasy of one good day's fishing in a month's time talking the rivers with the *Piscatore*, the *Viatore* admits he has matters to tend to, and yet he notes that "if business would give me leave, and that if it were fit, I could find in my heart to stay with you for ever."[4] It would appear that a "lecture of Flie-Fishing" is as seductive as the sport itself, and that "some of the one, and some of the other" creates the relational space between two like-minded individuals. Things may not be so heartfelt between anglers on the Yellowstone these days, and yet between hatches, between

business, for that matter, anglers will always listen to lectures, and give them, too.

The history of fly fishing in the western United States is perhaps not so richly constructed a textual space as it is in Great Britain, or the eastern portion of the country, for that matter. And yet there is little doubt that the thousands of anglers who flock to the waters of Yellowstone every summer feel as if they are participating in a rich culture of leisure---one which has been communicated to them no less seductively than the manner in which the promise of one good day fishing in the midst of a month's worth of lectures draws *Viatore* to the *Piscatore*, with the feeling that he could stay with him forever. They read books and magazines, they listen to stories. They ply the shops for new equipment and flies that will work, and they try to understand how they should make their appearance to one another, and to the natural world. They take time off, book flights and reservations in rustic, comfortable lodges, so that for a brief time they can feel like the Viatore, as if they could remain there forever.

The culture of fly fishing, I would like to argue, is bound up with notions of eternity; it is, in its stance toward the world, what the angler would like always and everywhere to be doing. And yet insofar as realistically, it is drawn against the finitude of time, it remains inscribed with the forms of desire, and somehow troubled with its relations to the natural world, the cyclicality of life, and its own ontological ties to the realm of Necessity, with the patterns of labor that haunt its history. There is a sense in which the tradition of fly fishing, so heavily codified in its practice, and laden with Anglo-European values and morality, can be understood as a sort of collective narrative which the culture tells about its beginnings and its "ends," and whose teleological impulse can be

understood with respect to attitudes of the human body, and its relation to others and to the natural world. The physicality of fly fishing, in other words, connotes a shifting ground of human effort which valorizes, and aesthetisizes what it perceives to be an activity that has cast off its fundamental ties to hunger, labor, and the reproduction of the life process-- those elements which fashion the ground of fishing as a type of bodily labor. The angler's appetite, effort, and productivity may indeed be retained as elements of the popular discourse on angling, but they are clearly sublimated, made harmless, so to speak, against a cultural understanding of leisure as it comes to be defined directly in opposition to "work."[5]

That this narrative comes to us as a story fraught with the ideology of redemption has become something of a given; the image of the fly fishing divine, the idea of communion with nature as a valid form of religious experience, and not least, a rash of tackle and tee-shirts inscribed with Norman Maclean's now clichéd reminiscence of growing up in a family where "there was no clear line between religion and fly fishing," all attest to a cultural attitude that most adherents of the sport would lay some claim to .[6] And yet clearly, too, the sport maintains a tenuous relation to the strict observance of Christian values; it's something of a commonplace for an angler to respond to a question of religious denomination to reply that he or she attends the "Church of the Gallatin River," or "Our Lady of the Henry's Fork," and so on, implying that, while fly fishing retains the terminology of Christian *redemption* in its popular discourse, it has come into a ground of contention with a notion of Christian *practice* that finds itself competing for hegemony with respect to forms of "rest," and the temporal space of the Sabbath.

Lewis B. France, whose *With Rod and Line in Colorado Waters* is something of an early classic on fly fishing in the west, suggested that this secular impulse, with its attendant feelings of guilt and fear of retribution, fashioned the ground of his initiation into the sport at the hands of one "John Doe," a familiar Sunday "pedestrian" in the village where France had his boyhood. In the last chapter of the book, entitled, "His Sermon," France admits to a childhood fascination with this individual, who, he notes, "Worked in blue cotton overalls six days a week... and walked on the seventh," or, testing the reader's grasp, "he walked on the first and worked on the other six days, as you please."[7] The figure of Mr. Doe is something of an enigma to the youth, who notes he "always walked with a cane," which was "for some time a mystery, he being an active man with no apparent use for support of that character" (*WRL*, 143). Just what there is beyond this cipher of the cane which draws the youth to the elder is unclear, and yet young France fantasizes about spending a Sabbath in his company, and he notes "I contemplated the bliss of walking with Mr. Doe with an eagerness hard to overcome," stating further that "I have sometimes felt that the fear of mere reproof, unaided by the respect in which I held the tender branches of the ..shellbark in our back yard, would not have prevented me from running away" (*WRL*, 144). However, having assured his caretaker that Doe would never be one to go hunting on a Sunday, France manages to secure permission to accompany him on his stroll, and on the following morning, Doe is found at the gate, where, the author notes, he "had his cane with him, and his shoes and linen presented their ordinary, unobjectionable appearance, as if defiant of criticism" (*WRL*, 146).

The course of their walk, of course, takes them to the bank of a stream that Doe pronounces is "a good place to fish," and it this point in the narrative, the text becomes a slow revelation of the motives behind the man's Sunday habits, as well as a discourse on the seductions of Sabbath fishing. Doe cuts a branch to make a pole for the boy to fish with, and forthwith begins to produce a covert arsenal of tackle from the "hidden recess of his coat," and then, more miraculously, France notes, "he proceeded to unscrew the ferrule of his cane, and lo! the inseparable walking stick was transformed into a rod; his own manufacture" (*WRL*, 147). The reader is given to understand that being seen with a cane on the Sabbath is much more acceptable than being seen with a rod; and yet the scene becomes more and more bound up with Christian terminology and imagery the more the reader becomes aware of just what is at stake here with respect to the recreation of a working man. France notes that it was not until lunch that "it dawned upon me to inquire of Mr. Doe if he did not think it wicked to fish on the Sabbath," and in the remarks that follow, Doe is more and more the advocate of an outdoor religion that justifies his behavior, though he tells the boy that "while he saw no impropriety in passing the Sabbath out-doors in the fresh air and sunlight, there were those who would be shocked at what they deemed a desecration" (*WRL*, 150). But the assertion goes nearly unnoticed, as Doe takes on more and more the figure of divine wisdom, who, as France notes, "felt responsible to a higher authority for his acts, and would render his accounts at the proper forum in due course of time" (*WRL*, 150). Finally, the scene undoes its original transformation, returns to hiding that which was drawn from the recesses, and what the reader comes to understand is that the Sabbath angler is involved in the maintenance of a

social contract, in which a few hour's fishing creates the conditions for living amongst others in the world. Comfortable in his ability to render his accounts at the proper time, Doe, "[m]eanwhile... propose[s] to follow the admonition of the great apostle: '*If it be possible*, as much as lieth in you, live peaceably with all men'" (*WRL*, 150). Prepared, perhaps, for the next six days, France notes that "[u]pon this he transformed his rod again into a walking stick," and "carefully stowed away the lines... we gathered up what had been vouchsafed to us and started for home" (*WRL*, 150). The secular impulse behind the idea of Sabbath angling, then, redeems the heart of the social contract; that it appropriates the terminology, the time, and space of Christian redemption, is something that may have to be accounted for, though Pauline doctrine tempers apprehension on this score.

If, indeed, angling is a means to living peaceably, and gracefully with others in this world, then it, too, is ultimately bound up with what Hannah Arendt refers to as the condition of "plurality" in this world, the "fact that men, not Man, live on the earth, and inhabit the world."[8] For Arendt, the condition of plurality is the essential dynamic in the world of social relations that constitute the life of the *polis*, or the social space in which individuals make their "appearances" to one another, and where each individual seeks distinction through great acts, and deeds, both spoken, as well as articulated through the body. "Action," in this sense, is constitutive of an *oeuvre* of deeds and acts which accrue over life of the individual, and which, as the space of remembrance and individual history, outlasts the ruination of the physical body.[9] I suggested earlier that the culture of fly fishing does, to a great extent, constitute a social

space of appearances, where being seen, and seeing others creates of the natural environment a sort of outdoors *polis*, in which deeds and exploits are on exhibit as individuals attempt to distinguish themselves in what often amounts to a silent discourse of physical activity. And yet, with respect to the culture of leisure, I would suggest that action is drawn to s pacific ethos of "counter-productivity," one which has to do with the expenditure of time as a visible component of leisure, and which has come to constitute a key element in our understanding of the concept, as articulated by Thorstein Veblen in his famous *Theory of the Leisure Class.* What he has to say with regard to the consumption of time in the pursuit of leisure is worth noting at length, since it speaks not only to a conception of the social, visible angler making an appearance in midstream, but also to the solitary angler, who, in creating the tale, the text, or the trophy, makes an appearance as well to his or her fellows:

> It has already been remarked that the term "leisure," as here used, does not connote indolence or quiescence. What it connotes is non-productive consumption of time. Time is consumed non-productively (1) from a sense of the unworthiness of productive work, and (2) as an evidence of pecuniary ability to afford a life of idleness. But the whole of the life of a gentleman of leisure is not spent before the eyes of spectators who are to be impressed with that spectacle of honorific leisure which in the ideal scheme makes up his life. For some part of the time his life is perforce withdrawn from the public eye, and of this portion which is spent in private the gentleman of leisure should, for the sake of his good name, be able to give a convincing account. He should find some means of putting in evidence the leisure that is not spent in the sight of the spectators.[10]

There are a number of things which have gone into the codification of non-productivity as a pattern of ethical practice which fly fishers, in one way or another, have come to understand limns the space in which they make their appearance to others. How that relates, perhaps, to an abiding sense of the "unworthiness of productive work," has become articulated in

the refrain one hears or sees often enough, that "the worst day fishing is better than a good day at work," a sentiment whose various incarnations are coeval with the notion of failure, and which can't it seems, be uttered seriously regardless of what sort of day one happens to be having. However, as Veblen makes clear, the wasteful expenditure of time, is essentially attested to in the form of the symbol, or evidence of deeds, actions, or exploits, and in this sense, a sort of minimal productivity, a sense of potency and skill that can be economized in the form of the limited catch, becomes the indice of leisure, and the means of appearance in the social arena of the sport.

Natty Bumppo, with who only Teddy Roosevelt might compete for the role of the national sportsman, is often distinguished by a sense of exploit bound up with a moral economy of self-restraint, and the limited catch. Though not a fly fisher, (this I don't understand) it seems worthwhile to consider briefly the remarkable angling scene which unfolds as a morally contested ground of practice in Cooper's *The Pioneers*, primarily because it so effectively places the images of radical productivity alongside the images of exploit and a moral economy of restraint. The scene, which unfolds on the dark banks of Lake Oswego, constructs the dual space of, on the one hand, a large social gathering of individuals who are out to secure a large catch of fish by casting nets, and on the other, the minimal gathering of Natty and Mohegan, who are drifting along the shore, as Natty is poised with a spear in his hand, looking to strike his supper. The group on shore has already hauled in a huge catch of fish from the lake, and it's clear to the reader that the productivity of fishing with a net has exceeded the limits of consumption, threatening to lay waste to the

resources of the lake. Cooper is at work here confusing the images of sport fishing with subsistence fishing, inasmuch as once the threshold of the community's ability to consume is crossed, this sort of angling quickly degenerates into a morally reprehensible form of sport. And yet, Natty's own intentions are equally complex; he, after all, is ostensibly looking for his meal, and yet, when the Sheriff tells him to "paddle in old boy, I'll give you a mess of fish that is fit to place before the governor," Natty rejects the offer.[11] His rejection is a moral imperative, however, since he replies "I eat of no man's wasty ways. I strike my spear into the eels, or the trout, when I crave the creaters, but I wouldn't be helping to such a sinful kind of fishing."[12] Instead, the Leatherstocking turns his canoe, and continuing his hunt in the light of the fires, pursues his supper. And yet if the mass netting of fish beyond anyone's conceivable needs degenerated into a perverse form of sport, Natty's subsistence practices become transitional as well, since what he accomplishes in the light of the fires accrues primarily to a notion of exploit, made sport by the very virtue of the economy of his cravings. Counter-productive to say the least, with respect to the practice of netting, he adroitly manages the single fish, in full-view, and "raising his prize by one of his fingers, and exhibiting it before his torch," he exclaims "I shall not strike another blow to-night."[13] The economy of appetite is here simply a gloss for an economy of practice, and Natty's exhibition devalues the productivity of a method whose roots are tied to subsistence, understood as a "worthless" ground of human action.

The history of angling in the west has its own moral ground of productivity, one which is tied to the very fecundity of the fisheries on the frontiers, and which could cast the aspersion of "worthless production" on

even the most lazy of anglers. Trout in the heavily fished waters of the east had become the prime targets of the angler who sought to limit a catch by virtue of an ideology of the worthy adversary, educated, selective, elusive--- in short, prone to disappointment. However, trout inhabiting the fisheries of the western regions could be too quick to the fly, sluggish in the fight, or just too plentiful to make a virtue of casting a line. T.S. Van Dyke, recounting for readers of *Forest and Stream*, was bold enough to suggest to his readers that the fishing in the Kern River was almost too good, and that after "[a]nother, and another, and another took the hook... it began to look as if the fish would spoil the sport by being too plenty."[14] The results of this early morning success have the angler on the verge of moral turpitude, and later in the afternoon, when the fishing is harder and Van Dyke manages to hook, skillfully, a larger fish, he has to confess his fear of failure, and his feeling that "I had been demoralized with too much easy trout fishing."[15] Before things fall apart, and before, perhaps, the reader makes their way to the fisheries of the west, Van Dyke asserts what can only be asserted in the face of such a dilemma: vouchsafe the public space of individual exploit, but waste more time, and practice some self-restraint. "The best way," he tells his readers, "is to let but one of the party handle the rod while the rest sit down and look on. When he makes a miss he must take a seat and wait until the rest have missed."[16] Patience is a virtue, in this instance of self-restraint, and Van Dyke was prepared to minister to those who might come west, and find their hands full.

There is another aspect of western angling which caused concern for early fly fishers in the west, and which, in some form or another, has

continued to the present. Fly fishers are always given to innovation in the construction of flies; the availability of materials, lack of success, aquatic insect life-- all these issues can lend themselves to a regional assortment of fly patterns that come to occupy the fly boxes of local anglers. however, the west has been the problematic site (more a projection than a truth, perhaps) of a type of fly whose success has been unquestioned, yet which bears no strict resemblance to a specific insect, and which seems to work primarily because of it's gaudiness, its exaggerated features, or even because the fish take it for some reprehensible reason, such as they've never seen it before. The "attractor" fly---colorful, bushy, often just big, has come to be understood in the context of the big waters of western rivers, where it works; and yet its history, too, is troubled by the specter of worthless production, and its infamous role in the success of western angling is not too difficult to trace.

Lewis France, who seems to have been something of a regular on the Middle Park sections of the Colorado and William's Fork rivers in north central Colorado, managed to concentrate his repulsion for the gaudy fly in the image of the fashionable, but naive angler, a figure whose role in angling literature has been a constant, to say the least. France draws heavily on this "young gentleman," whose skills are so poor that his utter lack of success begins to look like "a day spent in unsuccessful labor" (*WRL*, 33). The nexus of his failure, however, lies strewn upon "a broad-brimmed felt hat, the band of which was stuck full of flies of all sizes and a multitude of colors" (*WRL*, 33). One fly in particular, "a crimson fly that would have driven crazy any fish except a sucker," and with which the youngster "said he had whipped five miles of water... and could not get a

rise," is essentially the germ of a long, and sustained discourse on fly-fishing as an activity defined in all its attributes by discretion, moderation, and an abiding sense of humility (*WRL,* 33-34). The choice of a gaudy fly, in this instance, is a clear, and laborious substitute for the lack of skill evinced by the calm, subdued angler using an equally subdued fly; France's own favorite is the "gray hackle," a fly that is, "to the trout, what bread [is] to a civilized man" (*WRL*, 35). After providing the young angler with one of his own, he returns a virtual "body of evidence" attesting to what the youngster would have liked to kept private with regard to his exploits:

> [H]is drab pants were ruined, his rod had been shivered into kindling wood, his reel lay in a pool of the Grand [Colorado] twenty feet deep. He had cast that gray hackle with a brown body into that pool; it had been seized upon by a trout something "near a yard long;" the angler had succeeded in landing its head upon the rocks, then his rod gave way and he fell on the fish, rolled into the river, lost the remains of his tackle and his hat with the flies, and some other tenderfoot who happened providentially that way, had pulled him out by the collar (*WRL*, 34).

Exhausted by the lack of success, traumatized by too much success, the colorful angler fishing the attractor fly becomes the site for disgrace and physical calamity.[17]

Essentially, this issue turns on another complex notion of self-restraint that has been heavily codified in the tradition of fly fishing, and which made its way across the Atlantic virtually intact, in the form of a long-standing debate on the construction of flies, and how they ought to be fished. The "purest" form of the sport, as it has been handed down, and generally accepted as the image of the sport as it makes itself felt in popular culture, is casting the dry fly, drifting it on the surface of the water where the fish has to see it, and rise to take it. Rod makers such as

W.J. Cummins of Auckland, England, traded on the traditional appeal of hooking trout on a rise, and encouraged readers of *Forest and Stream* to "Use The Dry Fly This Season," adding that "to be successful and to get the most enjoyment out of this pleasant and scientific method of trout fishing, you must have Cummins' Rods and Tackle, which are of the HIGHEST POSSIBLE GRADE."[18] Trout, in general, take about 90% of their food in some other form than an insect floating on the surface of the water; there's little question, perhaps that its the most visually pleasant way to fly fish, more question as to whether it's the most difficult, but no question as to its status as an ethical way to practice the sport. As a result, the image of the dry-fly purist, sitting on the banks of a river waiting for a hatch of insects, while it may not be the subject of engravings or woodcuts, is a familiar enough image of self-restraint to anyone who spends much time on any of the more popular "dry-fly" rivers in the west.[19]

The dry fly, essentially, is born in an act of mimesis, so to speak; it seeks the faithful imitation of an aquatic insect on the surface of the water. It places currency in the skill of the individual to both observe, and faithfully recreate a feature of the natural world, but in so doing, inscribes a code of ethical practice which few could argue limits the angler's options on the water. The moral universe of fly fishing is buoyed in many ways by its material culture; a light fly rod and leader is about the furthest thing from a net that anyone could imagine. And the terminus, "those poetized hooks," that are dry flies, constitute an amazing redistribution of productive energy in the sport, insofar as great care and skill are expended in the aesthetic construction of an artifact, whose utility can

always be called into question.[20] And yet, the construction of the dry-fly as a faithful imitation, constitutes a valuable expenditure in its own right; Theodore Gordon, though he advocated that a "man who confines himself to fly-fishing on free-waters should be proficient with the wet, dry, and even the sunk fly," implied that the individual "creation of a few specific patterns "(most assuredly dry flies) kept the "fingers... busy in leisure hours," and he noted further that "there is great satisfaction in killing trout with the work of one's own hands."[21]

Gordon can make such a startling assertion of the dry fly as "killing tool," primarily because its efficacy is mediated by a codified, and moral structure which shrouds itself in an halo of aesthetic values, and which trades on a social economy of expenditure that makes a virtue of self-restraint and the limited catch. The practice of catch and release angling, and the designation of "fly fishing only" waters seem born more of necessity than the consummation of these values in the late 20th Century. And yet, "having a taste for the sport" more than for the fish, has always somehow implied that so long as one has time on one's hands, and plenty of it, the individual casting a fly can afford, as France suggested, to

> "let your discretion be your tutor,' and you will work out your own salvation more surely than by a library of directions, remembering this for an axiom, that: The true sportsman does not go downstream and afield for the mere love of killing something (*WRL*, 37).

Discretion is itself a posture toward the world, which draws on patience, the lack of desire, and an abiding sense that one's behavior is both ethical, and attractive by virtue of its understatement. The image of the angler on the water is ultimately aesthetic because the body itself is working out its own "salvation" in the form of a subdued gesture that is utterly

untroubled by craving, "wasty ways," and "sinful kinds of fishing." "[U]ntil man is redeemed," writes Norman Maclean, "he will always take a fly rod too far back, just as natural man always overswings with an ax or golf club and loses all his power somewhere in the air" (*ARR*, 3). Over-reaction is never the indice of leisure, as an idealization of how time might be spent when God is not deducting the hours.

Finally, even as the physical body, working out its own salvation in the form of the subdued, and graceful act is born in a rejection of desire, and craving, and the "mere love of killing," so too is the landscape of leisure often wrought out of the images of the pastoral as the culture's consummate landscape of discretion and grace. Writing toward the end of the nineteenth century about the New Mexico landscape as a sort of angler's paradise, John Carnifex advised his readers that, while much of the Rocky mountain west was still rough, inaccessible, and harsh in climate, "[o]ne river exists in the Rockies,

> where the angler can go with an assurance of catching plenty of trout, and gamy ones, and can carry on the fight without exhausting either his body, his patience, or his purse, and where he will find his vocation yield a return in health and amusement and improvement of mind and muscle quite beyond the pleasure of successful angling, which itself is enough to satisfy a reasonable man.[22]

The Pecos river, flowing through Northern New Mexico, is rife with images of the restful pastoral, and "takes its name--- the Spanish word for sheep--- from the vast pastures which lie along its lower courses," and where, Carnifex, notes, "thousands of flocks are herded night and day by sleepy Mexicans, as they have been for three centuries past."[23] I have little trouble, to be honest, imagining that Carnifex had the impression that three centuries of Hispanic culture was at work in the designs of

Providence in the slow, and patient construction of the model ground of fly casting. I have more trouble believing he saw any irony in his brief history of the "Catholic Friars," who, he thought, must have been "as eager to win souls and lay up treasures in heaven as their captains were on earth. Rejoiced at the discovery of this great harvest ripe for the Church," Carnifex notes, "they first fell upon their knees; then they fell upon the people."[24] For Carnifex, who worried that the pleasures of angling "must not cost too much in toil, weariness, or discomfort or else they will disappear under the shadows of these ills," this falling on the knees, this falling on the people, must have seemed an overstatement indeed.[25] Salvation, redemption, and the laying up of treasures in heaven could hardly have been so urgent a concern here in the pastoral landscape of leisure. Sleepy Mexicans, discrete fly casters, beautiful killing tools wafting through the air, this must have seemed like the active life of eternity.

---

[1] In *Gone Fishin'*, a collection of angling aphorisms, compiled by Marc Anello. (White Plains, NY: Peter Pauper Press, 1996).

[2] A permit to fish in the Park comes with a request that each angler fill out a log describing his or her experiences, dates, and waters fished, and species of fish caught. It's a good policy, and I only mean to draw on its rhetorical value here.

[3] See Paul Schullery, *American Fly Fishing: A History*. (New York: Nick Lyons Books, 1987), esp., "Angling Society," pp. 122-32, for a discussion of the growth of the conservation movement in angling, and its relation to exclusivity and privatization. Schullery's history is the broadest work on the history of the sport in America that I know of, and I am deeply indebted to it throughout this discussion.

[4] Charles Cotton, in *The Compleat Angler*. (Oxford: Oxford University Press, 1982), p. 271.

[5]See R.T. Allen, "Leisure: The Purpose of Life and the Nature of Philosophy," in Cyril Barrett and Tom Winnifrith, ed., *The Philosophy of Leisure*. (New York: St. Martin's

Press, 1989), pp. 20-33, for a discussion of the ideal of leisure as defined primarily in its opposition to work.

[6] Norman Maclean, *A River Runs Through It*. (New York: Pocket Books, 1992), p. 1. Hereafter, cited as (*ARR*). Though a "contemporary" novel by most standard's *A River Runs Through It*, is also a strong attempt to articulate the culture of angling in the American West of the early 20th century, and for that reason, I choose to consider it a sort of "source" work for this discussion.

[7] Lewis B. France, *With Rod and Line in Colorado Waters*. (Denver, CO: Chain, Hardy & Co., 1884), p. 143. Hereafter cited as (*WRL*).

[8] Hannah Arendt, *The Human Condition*. (Chicago: The University of Chicago Press, 1958), p. 7.

[9] See "Action," in Arendt, *The Human Condition*, pp. 175-247 for an extended discussion of what I have obviously generalized here.

[10] Thorstein Veblen, *The Theory of the Leisure Class: An Economic Study of Institutions*. (New York: The Modern Library, 1934), pp. 43-44.

[11] James Fenimore Cooper, *The Pioneers*. (New York: Bantam Books, 1993), p. 338.

[12] *ibid.*, p. 339.

[13] *ibid.*, p. 344.

[14] T.S. Van Dyke, "Among the Golden Trout," in *Forest and Stream* vol. XLVI #10 (March 7, 1896): 442-45, quoted p. 443.

[15] *ibid.*, p. 443.

[16] *ibid.*, p. 443.

[17] See Schullery, *American Fly Fishing*, esp. chapters 8 & 9, "Victorian Glories," and "Imitation's Forgotten Prophets," pp. 75-99 for a discussion of flies, to which I am indebted here, and in what follows.

[18] Advertisement, in *Forest and Stream* Vol. 9, #2, August 16, 1877.

[19] The popular stretch of the South Platte river near Deckers, Colorado, is still home to such figures, as well as the site of the Wigwam Club, a private angler's club ,of which John Monnett, in an introduction to Roger Hill's *Fly Fishing the South Platte River*, recounts the following anecdote of an excessively productive angler fishing flies beneath the surface, where trout are most likely to be feeding: "The story goes that Cliff Welch, a Wigwam member of good standing, was fishing the club property one fine August day in 1935. He was setting the hook in one large rainbow after another while his companions couldn't buy a strike. When asked what he was using, Welch produced a March Brown nymph tied to the end of a nine-foot leader with a split shot attached to the fly. Club officials were furious. They demanded that Welch give up the practice or be censored from the Wigwam Club, as using weight was not considered proper fly fishing." p. xvi, "Foreword," to Roger Hill, *Fly Fishing the South Platte River*. (Boulder, CO: Pruett Publishing Co., 1991).

[20] Anon. "Fishing Tackle and Things," in *Forest and Stream* vol. XLVI no. 2, March 14, 1876, p. 562.

[21] Theodore Gordon, "A Little Talk About the Angler's Flies," *Forest and Stream* Vol. LXXIX no. 20 (November 16, 1912): 626-27, pp. 627, and 626.

[22] John Carnifex, "Fishing along the Pecos River," *Outing* (July, 1891): 298-306, p. 300.

[23] *ibid.*, p. 300.

[24] *ibid.*, p. 300.

[25] *ibid.*, p. 300.

CHAPTER 5

"AT OUR MEANEST TASKS: REDEMPTION AND THE DOMESTIC ECONOMY IN WILLA CATHER'S *DEATH COMES FOR THE ARCHBISHOP*

Seated at his desk in the fading hours of daylight on Christmas day, 1851, Father Jean Marie Latour, Willa Cather's Bishop of Santa Fé, writes to his brother in a "fine, finished French script" of his home and his life on the frontiers of New Mexico Territory.[1] Just returned from a three thousand mile journey to Durango, and at ease in his new Episcopal residence, "an old adobe house, much out of repair, but with possibilities of comfort," Latour recreates in his letter the tranquil, domestic space which is a haven for the exile "thinking of home," yet making a life in the New World (*DCA*, 33, 36). For although the bishop's study is sparsely furnished, the residence is more than habitable, and the earthen hearth fills the room with "the delicious fragrance of... piñon logs," whose "highly aromatic, yet delicate" odor, he suggests, makes sacred the very fabric of the life of the home (*DCA*, 35). To be sure, neither Father Latour, nor his Vicar General Father Joseph Vaillant, are content to lead a life confined inside the walls of house or church; at Christmas dinner that evening, Latour recalls their missionary life in Ohio, and reminds Vaillant that he "used to chafe a little at [his] dear Sandusky and its comforts... to say that [he] would end a home-staying priest after all" (*DCA*, 39). And yet the sense of comfort and well being which permeates the spaces of "The Bishop Chez Lui" gathers its strength in the assurance of sanctuary at the end of the journey; the idea of *home*, in other words, lends an aura of grace to the missionary at rest in his wanderings, and makes of his hearth the instrument of blessing. "At our meanest tasks," he writes, "we have a

perpetual odor of incense about us;" the piñon's fragrant smoke lends sanctity and the atmosphere of ritual to the habits of domestic life, suggesting that Grace, security, and the instruments of redemption are near to hand in the most rudimentary acts, the most humble of spaces (*DCA*, 35).

There is a certain economy that shapes the lives of these priests who create not only a diocese, but a home in the New Mexico landscape; attentive to the needs of daily life on a harsh frontier, and yet clearly focused on the tasks of the spirit as both missionaries and men with a rich cultural heritage, experience unfolds as a mediation between the life of the body, and the life of the mind, between the urgency of the need to simply sustain one's self in the middle of a desert, and the desire that life itself take what is given it, and give it grace, beauty, and meaning. The two French Jesuits are no ascetics, and yet the comforts and pleasures of the home seem wrought from the stark simplicity of the landscape itself: the adobe structure, the gardens planted and tended wherever water can be found, the piñon, source of light and heat, even the furnishings of the home, 'heavy, and somewhat clumsy, but not unsightly," and "hewn from tree boles with the ax or hatchet" (*DCA*, 34). Given form, shaped by the hand of labor, the very *matter* of the New Mexico desert seems remade into the space and foundation of the human spirit, even as the Bishop's "theological books" rest on "thick planks" themselves "ax-dressed," and heavily inscribed with the marks of human effort.[2] The walls of the residence, "finished on the inside by the deft palms of Indian women," maintain the "irregular and intimate quality of things made entirely by the human hand," and are the source not only of a "reassuring depth and solidity" housing the physical body, but lend the patterns of labor to an

emotive/aesthetic life as well (*DCA*, 33). The adobe walls, "newly whitewashed in the Bishop's absence," provide both shelter and pleasant surroundings; lit by the piñon logs burning in the fireplace, the Bishop's study reflects "a rosy glow over the wavy surfaces, never quite evenly flat,... for the ruddy color of the clay underneath gave a warm tone to the lime wash" (*DCA*, 33-34).[3] Jean Latour, seated at his desk this Christmas evening, has the Nativity, and the promise of Christ's redemption to reflect upon as an exile, "happy," and "thinking of home" while he writes to his family in Auvergne; such is the source of comfort for a French Jesuit priest (*DCA*, 36). Reflecting on the light off the adobe walls, however, the New World missionary finds that the patterns of redemption blend readily with the exigencies of life; reassurance, comfort, shelter, these things take on a physical being in the act of giving human form to the landscape. In other words, the ideology of redemption, in those moments that reflect upon the works of the hands as components of a world made both safe and beautiful, finds occasion to set its scale to a model of domestic economy, and in doing so, valorizes those efforts, and products of effort, which sustain both life and spirit in a harsh environment.

Economy, sanctuary, sustenance, need--- *Death Comes for the Archbishop* ruminates persistently on the teleology of these concepts in Western thought, drawing together as it does a sparse, unyielding, "primitive," and yet spiritualized landscape, and the lives of individuals whose cultural heritage is made rich by Western art, architecture, *haute cuisine,* and the artifacts of Christian redemption, yet which is itself moderated by the implicit rejection of worldliness demanded by the Catholic church.[4] The locus of this meditation, of course, is the figure of Jean Latour, whose world-view and physical features seem both at home

and at odds with the New Mexico landscape, who sees both the Fall and the Redemption in the broad vistas of the desert and its indigenous cultures, and who is at once a churchman and frontier "businessman" who laments to his brother that "[w]e missionaries wear a frock coat and wide-brimmed hat all day, you know, and look like American traders" (*DCA*, 35). The Bishop is all too self-effacing in his comments, however, for Cather suggests that there is more to the man than his garments, and that he was "a priest in a thousand, one knew at a glance" (*DCA*, 18). To be sure, Latour emerges as an idealization of western man whose head, Cather notes, is "built for the seat of a fine intelligence," and whose physical features allow him to slough off the onus of a harsh frontier life---a priest whose manners, "even when alone in the desert," are "distinguished" (*DCA*, 18, 19). "[B]elow the fringed cuffs of the buckskin jacket" he wears, there is "a singular elegance about the hands," and "[e]verything show[s] him to be a man of gentle birth" (*DCA*, 19). So it is, then, that the Bishop embodies both the harshness of past and present on the New Mexico frontier, as well as the progressive, civilizing impulse equivocally registered in the image of the trader/missionary.[5]

There is little doubt, however, that Cather understood her priest to be a man who drew his essential character from the lineage of the refined European, and her remarks on the statue of Jean Baptiste Lamy, the historical model for Bishop Latour, are revealing:

> I never passed the life-size bronze of him which stands under a locust tree before the Cathedral in Santa Fé without wishing that I could learn more about a pioneer churchman who looked so well-bred and distinguished. In his pictures, one felt the same thing, something fearless and fine and very, very well-bred--- something that spoke of race. What I felt curious about was the daily life of such a man in a crude frontier society.[6]

The image of the Yankee constitutes only a vague perception of Anglo-European descent which is linked to mercantilism, military conquest, and the understanding that the seat of power in the Southwest has shifted from Mexico to Washington. Latour understands this, and writes that he intends "to help the officers at their task here," suggesting that he "can assist them more than they realize. The Church can do more than the Fort [Marcy] to make these poor Mexicans 'good Americans'" (*DCA*, 35-36). For Cather, however, the bronze statue standing before the church becomes the sight of an altogether different sort of vagueness, one in which ideals of race, gentle birth, good breeding, and fine features evoke a clear image of the pure European, yet whose figure remains "curious" precisely because it stands out against the idea of "crude frontier society" as starkly as the Midi-Romanesque cathedral stands out against the backdrop of adobe and desert. "All day," Latour tells his brother, "I am an American in speech and thought--- yes, in heart, too" (*DCA*, 35). In such moments, the ideology of race, gentle breeding, and distinction gives way to the exigencies of life and business in a frontier town. Such a character, an "American in speech and thought," can forego an historical consciousness in its attentiveness to the present. Yet within the walls of the home, the Bishop makes a nightly appeal to the history of Western culture and tradition, and in so doing, takes on the features that speak of "race" and refinement, those features, indeed, exhibited by the Bronze statue of Lamy Cather herself was so fascinated by. "What a pleasure to come home at night and put on my old cassock!" writes Latour, "I feel more like a priest then... and, for some reason, more like a Frenchman" (*DCA*, 35). The Catholic church, and the traditions of European culture, will of course be asserted, by the text's ending, through a much more solid

and imposing structure in the form of the Cathedral built of desert stone. For the time being, however, the domestic space is called upon to shelter both the Old World émigré and the historical consciousness that resides in such a figure. Warmed, housed, and fed by the New Mexico landscape, it remains somehow at odds with developments which are making themselves felt at large, during the "business" days of Santa Fé, of American policy, and the Catholic church's enigmatic role in that progressive, and largely Protestant enterprise.

The idea that domesticity has a specific economy about its spaces and activities, an economy that secures and safeguards the accumulation of tradition and culture in the historical consciousness, fashions the site of an intense meditation on cultural practices which both range across, and yet differ among, specific cultures as they interact in the Southwest. Foodways, in particular, are the nexus of a cultural exchange that is rarely benign, and which, again, tends to valorize European *haute cuisine* as an arena of traditional practices which, rather than simply recycling the patterns of daily life, accrue toward a refinement of the domestic space, and in so doing, make of the home an idealized "narrative" of cultural progress. The meal which completes the scene of the "Bishop Chez Lui," for example, becomes the occasion for a brief fantasy of culinary imperialism in the assurance that Father Vaillant, who has "sent away [the] Mexican woman" so that he can prepare Christmas dinner himself, will nonetheless "make a good cook of her in time," as well as an extended meditation on the idea of the well-prepared dish as cultural artifact, "the result of a constantly refined tradition" (*DCA*, 38).

These, of course, are the Bishop's comments on the "dark onion soup with croutons," which father Vaillant has prepared from the

materials to hand in the New Mexican desert (*DCA*, 38). The dish is a compromise, to be sure; Vaillant wonders "how can a man make a proper soup without leeks, that king of vegetables," and chafes at the thought of "eating onions forever," and worse, "dried beans and roots for the rest of our lives" (*DCA*, 39). It would be easy to overstate the sense of urgency in the priest's concerns about the fare of the missionary in the New World; after all, he is well-enough disposed toward a bottle of French wine "begged" from a wealthy hacienda owner, though, as he suggests, there is a "slight taste of the cork" (*DCA*, 40). "However," he adds, "it is quite good enough for missionaries," if questionable for Frenchmen. And yet Vaillant's fears are far-reaching enough, with the evocation of a temporal progression which lasts "forever," for the "rest of our lives," yet which is counter-inscribed with a regressive imagery leading back in time and culture to a subsistence lifestyle made up of onions, dried beans, and finally, roots. Complicated as it is by the missionary's understanding of a pious asceticism which mediates the Epicurean tastes of the Frenchman, Vaillant's appeal to Latour that "[s]urely we must find time to make a garden" nevertheless suggests a concern that the New Mexico desert cannot, in itself, yield the materials of civilized progress, any more than the "rich Mexicans," though "[t]hey know its worth," "know how to keep.. properly" a bottle of French wine (*DCA*, 39, 40). Both the production, and the maintenance of *haute cuisine*, of material culture, remain problematic with regard to the future of the frontier.

If Father Vaillant, however, seems overwrought in his resignation that it will always be the missionary's fate "to plant where another shall reap," as if he can never expect to catch up to his own cultural heritage precisely because his ground is by nature uncultivated, the Bishop is less

intent on the compromise which substitutes one material for another in the soup, so long as the effort expended in its production maintains the purity of tradition, knowledge, and cultural specificity which accrues and resides in the historical consciousness (*DCA*, 39). "Think of it, *Blanchet*," he tells Vaillant, "in all this vast country between the Mississippi and the Pacific Ocean, there is probably not another human being who could make a soup like this" (*DCA*, 38). For Latour, the *soupière* is the repository of a collective human effort which can make itself felt as a matter of taste, and which, moreover, can redeem the present moment which Father Vaillant understands to be fraught with compromise, and bleak hopes for the future. For although Vaillant is "losing no time in reflection," having "tucked a napkin over the front of his cassock" and begun his meal in earnest, Latour's critical tasting of the soup becomes a deep moment of historical awareness, one which recreates the present as the site of refined and domesticated human effort drawn from the past, and economized in the figure of the cook. (*DCA*, 38). "I am not deprecating your individual talent, Joseph," says Latour to his friend (they are speaking French on Christmas day), "but, when one thinks of it, a soup like this is not the work of one man" (*DCA*, 38). Far from deprecating individual talent, the Bishop understands that, as the soup is the artifact of French culture, so is Vaillant the product of labor expended as part of "a constantly refined tradition," and ultimately, the vehicle for cultural transmission, if he can indeed make a good cook of the Mexican woman sent away for the evening (*DCA*, 38). "There are nearly a thousand years of history in this soup," he reflects, and his sense of the comfortable past of French *haute cuisine* is nearly as far reaching as is Vaillant's fearful insight into a future fraught with scarcity and subsistence living (*DCA*, 38). The Nativity may indeed

be the *event*, the hope of redemption in the global wilderness of the Christian narrative; on the New Mexico frontier, however, Christmas dinner shared by two French Jesuits redeems the space of the home through accrual, domestic economy, and the figure of a cook whose individual talent is built of event upon event, labor upon labor, and whose works, finally, are judged by the man of taste, reflection, and refinement.[7]

The discourse on foodways, then, in *Death Comes for the Archbishop*, carries within it a civilizing impetus whose teleology charts a course of progress from dried beans and roots whose sole value lies in the maintenance of life, to that moment in which the refined, historical dish is the occasion for contemplation; there is far less urgency in the Bishop's eating habits than in his friend Vaillant's, who seems to need to constantly replenish "the driving power of a dozen men in his poorly-built body" (*DCA*, 38). Latour, whose meditations had earlier allowed him to defer his thirst while lost in the desert by concentrating on that "cry, wrung from his Savior on the Cross, '*J'ai soif!*'" appears to understand his own physical needs as a locus for contemplating sheer ideas and essences whose true import are only inscribed in the body in the most literal manner (*DCA*, 20). Beyond the physical sufferings of God made man, beyond his own suffering, lay the redemptive idea of the Passion itself, and there in the desert, "[e]mpowered by long training:

> the young priest blotted himself out of his own consciousness and meditated upon the anguish of his Lord. The Passion of Jesus became for him the only reality; the need of his own body was but a part of that conception (*DCA*, 20).

For the reflective, contemplative individual, then, the physical body which bears the stamp of hunger, of thirst, of "need," is but a conceptualization of the necessity inherent in the ideology of Christian

Redemption, or in that of the capacity French culture to domesticate the broad diocese of Santa Fé.[8] The deferral of body to spirit charts a narrative of progress which valorizes both the Christian world-view, and that of European high culture. Indeed, during the succeeding courses of the meal, Latour's meditations seem to slowly draw Vaillant's concerns away from the meal at hand, and away from the labor of the missionary, toward Vaillant's apparently lost "resolve to lead a life of contemplation," expressed while still a young seminarian in France (*DCA*, 41). The tasks which lie ahead of the "wiry little priest who was to carry the cross into territories yet unknown and unnamed," and "who would wear down mules and horses and scouts and stagedrivers" require a voracious appetite (*DCA*, 41). And yet the mention of the *vita contemplativa* draws together both Christian and French perspectives in a soteriological fantasy of the body at rest, and Cather notes that at the thought, a "light leaped into father Joseph's homely face" (*DCA*, 41). "I have not yet renounced that hope," he states, and adds:

> One day you will release me, and I will return to some religious house in France, and end my days in devotion to the Holy Mother. For the time being, it is my destiny to serve her in action (*DCA*, 41).

And yet the vastness of the diocese, the plea that New Mexico be "far enough" in the missionary's travels breaks the reverie, and, "making haste to change the subject," Vaillant serves another course. "A bean salad was the best I could do for you," he tells his superior, "but with onion, and just a suspicion of salt pork, it is not so bad" (*DCA*, 41).[9]

Beans and salt pork are the more stable fare of the missionary in the New Mexican desert, not merely at home, but abroad, where the two priests often find themselves as guests at the tables of Mexican *rancheros*, priests, and occasionally, Native Americans. The image of the traveling

missionary binds Old World to New in *Death Comes for the Archbishop.* In the text's opening scene in Rome, the Irish born Father Ferrand, "a man of wide wanderings and notable achievement in the New World," is described by Cather as "an Odysseus of the Church," and Bishop Latour discusses "a fierce little equestrian figure" with a young boy at Agua Secreta, a figure he comes to understand represents the missionary St. Jaques, but who is known as Santiago in Hispanic culture (*DCA*, 4, 28). Latour reflects "[h]e was a missionary, like me," and tells the boy "[i]n our country here carries a staff and a wallet--- but here he would need a horse, surely. (*DCA*, 28). Hermione Lee states that "*Death Comes for the Archbishop* is marked out and moved along by the mission priests' enormous, adventurous, arduous journeys, the novel's substitutes for a sustained plot," and which seem to imbue the text with the quality of epic.[10] Yet those journeys are almost always the occasions for a close examination of architectural spaces: mission churches, Native American pueblos, natural shelters, and again, most often, the space of the household, as the site of specific cultural practices and the values they represent. Travel, more than anything, in *Death Comes for the Archbishop*, furnishes Cather the occasion to utilize the trope of the supplicant in need of food and shelter, as a means of examining cross-cultural perspectives on physical spaces and domestic habits.

If Christmas dinner in the Episcopal residence had offered reassurance in the form of *haute cuisine* as a sort of repository of French culture, tradition, and historical continuity, meals taken with inhabitants of the diocese at large are nowhere near so clear in their implications. Cross-cultural contact which revolves around foodways can be confusing, and while the priests are given to making special requests with regard to

the preparation of their meals, their hosts are often at odds to understand precisely what motivates tastes and preferences which appear out of step with prevailing local habits. Josepha, who prepares the Bishop's meal at Agua Secreta, and who appears as a model of the devout Catholic woman, seems unable to distinguish between European tastes and religious devotion, and when he makes a request to have a portion of goat served to him without chili, Cather states that "the girl inquired whether it was more pious to eat it like that" (*DCA*, 30). Latour sets the issue aright, explaining that a cultural preference, rather than a religious dictate, is the heart of the matter, since "Frenchmen, as a rule," he tells her, "do not like high seasoning" (*DCA*, 30). The Bishop is worried about his influence on the girl, and he makes the explanation "lest she should here-after deprive herself of her favorite condiment" (*DCA*, 30). However, the reader is left to consider the extent to which what Frenchmen prefer "as a rule," will ultimately guide the young girl's sense of domestic economy; Catholicism and nationality are so concentrated within the image of the French Jesuit, that it seems difficult, after all, to separate piety from taste. With regard to a young Hispanic girl in charge of a household, a girl whose devotion will make itself visible in pious imitation, preferences in taste can hardly avoid the implications of a morality inherent in issues of chili and high seasoning.

Cultural attitudes toward food preparation can be much more openly contested in *Death Comes for the Archbishop*, however. Hispanic households such as those seen during Father Vaillant's visit to the *rancho* of Manuel Lujon, are not always prime targets for religious or cultural influence in the text; in such households, where open concubinage, lax religious devotion, and unbaptised children suggest a lack of concern with

the Catholic faith in the New World, a careless domestic economy tends to rule as well, even when buoyed by wealth, and prosperity.[11] Father Vaillant's intentions, however, are to set the house right with God by performing marriages and baptisms, even though Lujon tells him upon arrival "[t]here is no hurry," and suggests "[a] little wine, a little bread, coffee, repose--- and then the ceremonies" (*DCA*, 54-55). Vaillant, however, rejects the prospect of a meal and repose until afterwards, and essentially turns his voracious appetite towards marrying couples "with great dispatch," losing no more time in reflection with religious ceremony than he had with his Christmas dinner (*DCA*, 56). However, two women hastily sent in to scrub the floor of the *sala* "doubt if all this marrying will make them any better," and their concerns that the household is out of order with births coming before marriages echo Vaillant's demands to perform "the marriages first, the baptisms afterward," adding further "that order is but Christian" (*DCA*, 54-55).

A lack of concern on the part of Manuel Lujon with regard to household order carries over directly into a discussion of dining habits at the *rancho*. Father Vaillant, whose dispatch with the marriages seems to have been driven by the fact "that he had been fasting since an early breakfast," inquires at what hour dinner is served, a remark to which Lujon replies "[w]e eat when it is ready--- a little after sunset, usually" (*DCA*, 56). Meals at the *rancho* are apparently served in tune to patterns of male labor and natural cycles; the behavior of males in the field, in other words, translates directly to a domestic economy in which the preparation of food is problematized by the lack of a precise moment in which to serve it. When Lujon tells Vaillant that he has "had a young lamb killed for your Reverence," he pays his respects to the Catholic

priest, yet fails to understand that the French cook has questions about method as well as matter, and he hardly expects Vaillant's further inquiry as to "how will it be cooked?" (*DCA*, 57). Lujon, anything but reflective with regard to the tradition, history, or taste that might accrue in the preparation of a good dish, offers his best guess as to how the lamb will be prepared, though he seems hardly to have considered the issue at all. Cather notes that Señor Lujon shrugged," and simply replied "Cooked? Why they put it in a pot with chili, and some onions, I suppose" (*DCA*, 57). Such a dish is prepared, of course, so as to be open ended with regard to its serving, and it relates to a specific domestic economy which values food primarily as a means for the reproduction of labor on the *rancho*, and there is little indication Lujon is willing to see things any other way. It is as if the thousand years of history gone into a well-prepared, and timely served soup, gives way to the hours a stew can sit at the hearth, and maintain its value as one of the means to wealth on the *rancho*. Food, then, is never really the occasion for repose or reflection, but spurs activity that makes itself felt beyond the limited sphere of the domestic space. In response to Vaillant's request to be allowed to go into the kitchen and cook [his] portion his own way," Cather states "Lujon waved his hand. 'My house is yours, Padre,'" he tells him, '[i]nto the kitchen I never go--- too many women" (*DCA*, 57).

Vaillant's journey into the kitchen is, in its own way, as epic as any of his long wanderings in the desert. Certainly, it ruminates heavily on national tradition, meditates on the idea of the heroic cook, and places him in opposition to an icon of Southwestern foodways who one suspects only yields her kitchen to the priest out of respect for his piety, rather than out of deference to his skills as a cook. Rosa, whose repository of

tradition is a "kettle from which issued the savour of cooking mutton fat, all too familiar to Father Joseph," has charge of the kitchen, yet she becomes quickly the unwitting servant in a cultural confrontation of culinary value which will make itself felt at the dinner table of Manuel Lujon (*DCA*, 57). Vaillant, who has discovered "a half sheep hanging outside the door, covered with a bloody sack," tells the woman than he "mean[s] to roast the hind leg," and immediately a sense of urgency arises which places time at a premium, directly in opposition to the slow, careless simmer of the mutton stew over the fire. "But Padre," she tells him, "I baked before the marriages. The oven is almost cold. It will take an hour to heat it, and it is only two hours to supper" (*DCA*, 57). She sees the task as daunting, and yet she misunderstands precisely because of the gulf which separates what may as well be altogether different concepts of time in contrasting world-views. "Very well," he tells her, "I can cook my roast in an hour," adding up, and drawing together both preparation time and serving time in a dish which will go directly from oven to plate (*DCA*, 57). Rosa is scandalized at the thought, which calls forth her peculiar devotion in the face of the priest. "Mother of God, Padre" she exclaims "the blood will not be dried in it" (*DCA*, 57). The rest of the scene plays itself out at the dinner table, and Cather notes that as "the Padre carved his roast at the supper-table, the serving girls... looked with horror at the delicate stream of pink juice that followed the knife" (*DCA*, 57-58). There is hardly any suspicion that Cather ruminates here on the hope of cultural transmission, and she notes that Manuel Lujon took a slice for politeness, but he did not eat it. Father Vaillant had his *gigot* all to himself" (*DCA*, 58).

The imagery of pious foodways ought to make itself felt heavily in this scene, evoking so blatantly as it does the concepts of sacrifice, the lamb, and the blood of the lamb. Yet the ground of cultural contention seems to diffuse the potential for meaning in the image of the roast. Then, too, the reader understands at this point that Vaillant loses no time for reflection in his meals, and the scene is much more compressed than is the earlier one at the Episcopal residence. However, while French cuisine is not the occasion for contemplation here at the *rancho* of Manuel Lujon, it retains both a currency of national tradition and preferences, if radically economized in the image of a one hour roast. History goes into the making of taste, after all, but more so, Father Vaillant's voracious appetite and timely served meal offers an alternative mode of production in the text, one in which the preparation of food comes to revolve much more strongly around the idea of the orderly household, and a specific brand of "domestic time." Later in the text, reflecting warmly on his friend's fondness for "good eating and drinking," Latour marvels that:

> Time and again the Bishop had seen a good dinner, a bottle of claret, transformed into spiritual energy under his very eyes. From a little feast that would make other men heavy, and desirous of repose, Father Vaillant would rise up revived, and work for ten or twelve hours with that ardour and thoroughness which accomplished such lasting results. (*DCA*, 226)

The two priests are, throughout the text, viewed in opposition to one another; Latour, the reflective and contemplative, Vaillant the active, laboring missionary at large in the diocese. At the *rancho* of Manuel Lujon, no one is reflecting too strongly on a meal; rather, food becomes a pathway to confront one sort of productivity with another, and if Vaillant eats as well as Manuel Lujon's men, it should be noted that what accrues to the priest is a "spiritual energy," economized in a religious/nationalist

gigot sanctified by French taste and the ideology of sacrifice. Such energy translates directly to the labor of making sacred the home, putting the house in order, and performing ceremony with dispatch. Stewed mutton buoys the labor of the fields, and makes of the home a careless, untimely world of affairs. However, it is Father Vaillant who remains the scandal in the scene, whose manners appear barbarous and uncivilized, and Cather seems to mitigate the ideal of cultural influence where spiritual, and material productivity too strongly confront one another at the meeting ground of domestic and business space, that space of the *portales*.

Evelyn Helmick Hively has argued that Cather's novels can be read as an extended national discourse modeled on Vico's cyclical view of historical process. Summarizing, she states that "[e]ach nation's development goes through succeeding ages of gods, heroes, and men from a state of barbaric simplicity to one of reflective intelligence."[12] Given Cather's turn toward Episcopalian religion previous to her writing of *Death Comes for the Archbishop*, as well as the pervasiveness of the text's Christian world-view, it seems unlikely that the "reflective intellect" concentrated in the figure of the Archbishop is so finely tuned to Vichian progress as a secularization of the human need for understanding and self-realization.[13] However, it remains clear that the home space in *Death Comes for the Archbishop* contends with specific world-views which meditate on cultural hegemony in the American Southwest, and which, moreover, construct a developmental fantasy in which Cather's Europeanization of national progress appears more domesticated, and benign than might be born out in the realities of Manifest Destiny. The shift from barbaric simplicity to reflective intelligence, so long as it remains localized in the images of domestic economy, remains an ideology

of progress largely untroubled by outright violence and usurpation.[14] There continues to be a strong tension, nevertheless, wherever the domestic space mediates cross-cultural values, and while it seems clear that European civilization operates under a halo of cultural imperialism seen as the cultivation of proper foodways, household order, manners, etc., Cather suggests that relations of power do not always coincide with a moral domestic economy. For the Catholic church, as a figure of authority in the Southwest, remains a juxtaposition of Old and New Orders; the question of national progress, which for the moment seems to rest with the fate of the Church in New Mexico Territory, seems threatened at moments in the text by an Hispanic priesthood whose hold on the territory is linked to a sensual religion, vice, and the radical appropriation of labor and resources in a "bad faith" domestic economy.

The central figure of Hispanic authority in *Death Comes for the Archbishop,* is Padre Martinez, of Taos, who, like Latour, is a remarkable in feature and mien, "not a man one would easily forget" (*DCA*, 140). Yet an unlikely study for a statue, his face seems to stand in opposition to that which speaks of refinement and gentle-birth, though it draws heavily on a concept of race which stands in contradistinction to that of the Bishop: "the full-cheeked, richly coloured, egg-shaped Spanish face" of Martinez commands attention, and Latour ruminates after their first meeting that it "was so unusual that he would be glad to see it again" (*DCA*, 140). Perched on "broad high shoulders," which "were like a bull buffalo's," and "set defiantly on a thick neck," Padre Martinez' face seems to resist the capacity for contemplation, with it's "full, florid cheeks, --- not blank areas of smooth flesh, as in Anglo-Saxon faces, but full of muscular activity, as quick to change with feeling as any of his features" (*DCA*, 140). His

mouth is wrought in tune to a sensual, bodily nature, "the very assertion of violent, uncurbed passions and tyrannical self-will; the full lips thrust out and taut, like the flesh of animals distended by fear or desire" (*DCA*, 140-41). As the Bishop meditates on the figure of Martinez as he approaches on horseback for a second meeting, it is clear that the old priest's clothes are remarkable as well--- "in buckskin breeches, high boots and silver spurs, a wide Mexican hat on his head, and a great black cape wound about his shoulders," he looks something like the statue of the equestrian missionary Santiago Latour had seen at Agua Secreta" (*DCA*, 141). To be sure, Martinez approaches as an icon of an Hispanic culture whose presence in the American Southwest had constructed its own sense of tradition and history; the martyrdom of the Spanish Franciscan friars, the Adobe churches scattered throughout the countryside, (of which Cather said "no record of them could be as real as they are themselves. They are their own story.") centuries of *ranchero* culture, all come together in the "picturesque and impressive" image of the Padre of Taos as he meets the Bishop outside of town (*DCA*, 141).[15] And yet Latour, who views this aspect of Hispanic culture concentrated in the figure of one individual with whom he must contend, sees, for the most part, the embodiment of "lawless personal power," and in doing so, seems to disperse, and ultimately relegate to the past, whatever might have accrued of race, tradition, and religious value, in the representative image of the Hispanic priest (*DCA*, 141). The history of Hispanic Catholicism written so strongly on the body, it appears, is indeed "already like something picturesque and impressive, but really impotent," perhaps the site of nostalgia, but hardly an aspect of the future in New Mexico Territory (*DCA*, 141).[16]

As Latour and Martinez ride into Taos, it becomes evident that whatever is extravagant, sensual, and overstated in the figure and fashions of the Padre of Taos, extends to the parish as well, and the Bishop endures a throng of "men and women," whom Cather notes "snatched for his hand to kiss the Episcopal ring" (*DCA*, 142). Latour is both patronizing and accepting of these gestures, which "[i]n his own country would have been highly distasteful," though he understands that:

> Here, these demonstrations seemed a part of the high colour that was in the landscape and gardens, in the flaming cactus and the gaudily decorated Altars,--- in the agonized Christs and dolorous Virgins and the very human figures of the saints. He had already learned that with this people religion was necessarily theatrical. (*DCA*, 142).

Drawing on the landscape more fully than the aesthetic traditions of European Catholicism, the New Mexico church is a naturalization the Christian redemptive schema, in which the Passion radically relocates its essence in the idea of God made man, in which piety, to a great extent, loses its capacity for reflection and contemplation, and remains tied to the bodily representations of "anguished" and "dolorous" human nature. The agitation of the body, the "necessarily theatrical" behavior of the Hispanic Catholic, can indeed be the source of a religious aesthetic, and even a traditional architecture; Cather herself suggested that the mission churches of the Southwest had left her with a strong impression:

> The old mission churches, even those which were abandoned and in ruins, had a moving reality about them; the hand carved beams and joists, the utterly unconventional frescoes, the countless fanciful figures of the saints, no two of them alike, seemed a direct expression of some very real and lively human feeling. They were all fresh, individual, first-hand.[17]

But again, such sites remain tied so strongly to the landscape that if anything, they participate in a view of Hispanic culture which is primitivized, and ultimately colonized by a more refined aesthetic gaze.

"In lonely, sombre villages in mountains" Cather suggested, "the church decorations were sombre, the martyrdoms bloodier... the figure of Death more terrifying," while in "warm gentle valleys everything about the churches was milder."[18] Cather thrusts Hispanic culture into the polarities of landscape art with this statement, leaving its religious world-view to wander between a wilderness sublime and a pastoral picturesque. Hispanic Catholicism remains the "subject" of religious art, and its construction as a realm of nature rather than of civilization is what draws it away from a more refined Christian redemptive scheme.

Again, however, Cather suggests that the domestic spaces which inhabit the landscape of *Death Comes for the Archbishop* are economized spaces in which to meditate on broad cultural values, and Latour's visit with Martinez moves "quickly across the grey plain" between Los Ranchos and "Taos itself, to the priest's house," ultimately the privileged site of the confrontation between Old, and New Orders in the text. The house itself is a model of disarray; entering the Padre's study, the Bishop is surprised to find Martinez' nephew "lying on the floor, fast asleep," in an exaggerated posture of bodily repose, stripped of the value of conscious thought and reflection (*DCA*, 142). Though he is "told to consider the house his own," Cather suggests that "He had no wish to. The disorder was almost more than his fastidious taste could bear" (*DCA*, 143). The entire image of the study is a direct oppositional image to that of Latour's, the "table sprinkled with snuff, and piled so high with books that they almost hid the crucifix hanging behind it" (*DCA*, 143). Where in Santa Fé, the bishops religious artifacts, and study materials had been sheltered, and made secure by the deft palms of Native American women, here, the books and the floors were deep in the dust of spring sand-storms" (*DCA*,

143). "Yet the place seemed over-run by serving women, young an old," Cather notes, although the most prescient image of female activity is a "bunch of woman's hair" the Bishop finds in his room, "indolently tossed into a corner when some slovenly female toilet was made there," a discovery which, Cather states, "annoyed the Bishop exceedingly" (*DCA*, 144, 149). Nature, and human nature, has gotten the better of this household. An attention to worldly affairs and business had indeed problematized the space of Manuel Lujon's *rancho*; there too, Catholicism held form as a matter of observance, and deference to religious authority. But here at the source of Hispanic Catholicism, where the embodiment of its essential nature is imaged in the form of the Padre and his disorderly household, not even the appeal to the picturesque, or sublime landscape of a Southwestern church can redeem this counter-statement of the European world-view.[19]

However, Padre Martinez is prepared to lay claim to nature as a redemptive arena for the individual, who, if not exactly in search of the finer points of Christian Grace in the midst of worldly affairs, might seek some understanding of the nature of sin, and the fundamental nature of the human condition. Martinez, who is well schooled in his own fashion, speaks freely, and during a dinner in which the "bishop found the food poor enough, despite the many cooks," he asks Latour "flatly if he considered celibacy an essential condition of the priest's vocation" (*DCA*, 145). The Bishop's response is surprisingly unreflective, Cather noting that he "replied merely that this question had been thrashed out many centuries ago and decided once for all" (*DCA*, 145). There is something in the dogmatic assertion, however, in which the denial of the eroticized body comes as a foregone conclusion to Latour; in the desert he had meditated

on the Passion of Christ, on physical anguish and suffering, and had discovered the means to the deferral of his own bodily needs. Yet desire, postulated as the erotic, as the heavily naturalized sphere of human reproduction, seems too strongly tied to the biological imperative to admit of reflection or conceptualization, where Latour is concerned, and rather than the possibility of sexuality being taken up as an aspect of Christian redemption, the question which "had been thrashed out many centuries ago" is a flat rejection of the endless cyclicality of the life process seen from this perspective.

Domestic economy, in the household of the Padre, resists altogether the imagery of celibacy, creating in its place a "religious order" over-run by the endless reproduction of books piled everywhere, cooks and serving women, as well as a devotional posture which is itself bound up with sexuality and desire. Martinez' young nephew Lucero, whose preparations for taking the orders are scandalized by "a titter of feminine laughter from the windows across the court, in response Padre's admonitions that he "must have been studying by candlelight," is a heavily sensualized individual whose gluttony at the table is coeval with his sexual desire. (*DCA*, 143). Latour watches him at the table, where, "[w]hen his attention left his plate for a moment, it was fixed in the same greedy way upon the girl who served the table," suggesting that the young man was "always stupefied by one form of sensual disturbance or another" (*DCA*, 145). The house itself, is over-run by "large yellow cats, with full soft fur" which the Bishop finds sleeping "in the window sills... on the well-curb in the *patio*," and "at the supper-table, where their master fed them carelessly from his plate" (*DCA*, 145). Most notable however, is Latour's sense of a chaotic and reproductive household staff, the endless

serving women and cooks who seem to breed like the cats. The Bishop's Episcopal residence had been the site of a domestic economy buoyed by the "deft palms" of Native American women whose absence is the mark of labor completed, and the radically economized figure of the singular cook drawing upon history in the production of a meal. Everywhere in the villa of Padre Martinez, household order seems undone by an overly reproductive domestic economy, driven by the sensual theology of young Lucero, whose stupefaction by sensual disturbance undoes the space of reflection and contemplation.

Martinez suggests that Latour's acceptance of religious dogma resides in a nationalist-cultural world-view as much as a Catholic one, and noting that "[n]othing is decided once and for all," he tells the Bishop "[c]elibacy may be all very well for the French clergy, but not for ours" (*DCA*, 145). To be sure, Martinez' whole understanding of Latour's doubtful claim to authority in New Mexico rests on the assumption that the European church has lost its ability to understand the nature of the relationship between sin and redemption in its cold denial of worldliness. "Celibate priests lose their perceptions," he states:

> No priest can experience repentance and forgiveness of sin unless he himself falls into sin. Since concupiscence is the most common form of temptation, it is better for him to know something about it. The soul cannot be humbled by fasts and prayer; it must be broken by mortal sin to experience forgiveness of sin and rise to a state of grace. Otherwise, religion is nothing but dead logic. (*DCA*, 146).

For Martinez, then, piety engages in a sensual activism that resists the static image inherent in the posture of devotion and contemplation. Arguing that not only has the church been cut off from Roman authority, but from the remnants of the "Church the Franciscan Fathers planted here," the Padre isolates his parishioners by stating that the New

Mexican Church "is the second growth, and is indigenous" (*DCA*, 146-47). The organic metaphor pits a religious naturalism against the "dead arm of the European Church," and he loses no time asserting to the Bishop that the question of Latour's authority lies in the fact that "Nature has got the start of you here.... Our people are the most devout left in the world. If you blast their faith by European formalities, they will become infidels and profligates" (*DCA*, 146-47). Devout human nature, it seems, animates sin, desire, and an indigenous cultural growth as the accouterments of a faith that can be "blasted" by the lifeless posture of formal piety. And yet there is no mistaking the overturning of the redemptive scheme inherent in this view. The primitivism evoked by the Padre threatens to reverse the ideology of grace constructed in the blending of European religion and high culture, arguing as it does that a pervasive mysticism is the true source of the Church in New Mexico. "You are a young man, he tells the Bishop:

> And you know nothing about Indians or Mexicans. If you try to introduce European civilization here and change our old ways, to interfere with the secret dances of the Indians, let us say, or abolish the bloody rites of the Penitentes, I foretell an early death for you.... You are among barbarous people, My Frenchman, between two savage races. The dark things forbidden by your Church are a part of Indian religion. You cannot introduce French fashions here. (*DCA*, 147).

The introduction of European civilization and "French fashions" is an enterprise made problematic, Martinez' suggests, because of Latour's lack of understanding of indigenous Southwestern cultures. And yet the "savagery" of Hispanic Catholicism, for all the Padre's assertions that "Rome has no authority here," remains tied to the Christian schema, largely bound up with the imagery of the sacrifice at Calvary, the Passion of Christ which Latour himself had drawn upon in the desert, if less spectacularly than the Penitentes (*DCA*, 146). However, Martinez is more

to the mark in his assertion that Latour has little knowledge of Indians, or those "dark things" forbidden by the European Church which Martinez states constitute a part of Indian religion. There is little mistaking the Bishop's deference to Native American cultures in *Death Comes for the Archbishop*; his enduring friendship with the Navajo Eusabio, his reflections on the events at Canyon De Chelly, and notably, his relationship with the Pecos guide Jacinto, all suggest an attitude of respect with regard to Native American cultures. Guy Reynolds has argued that the more genial moments of cross-cultural contact in Cather's works serve both as a model of cultural transmission, as well as a re-establishment of one's own sense of cultural value and tradition. Placing "affectionate regard for one's homeland and interest in the foreign alongside one another," he suggests, constructs "a model of cultural receptivity where feelings of national, if not parochial, loyalty [are] reinforced by exposure to cultural difference." [20]

Latour's interest in the "foreign" nature of Native American culture is a constant subject of meditation during his missionary travels throughout the diocese. Genuine moments of cultural exchange, however, are most often mediated through his relationship with the young guide Jacinto, whose pueblo of Pecos, Cather notes, "was dying-out; infant mortality was heavy, and the young couples did not reproduce freely" (*DCA*, 123). The communal space of the pueblo, "its rows of little houses all alike, and all built together," is something of a celibate space in its own right; traumatized by Anglo-European contact, and marred by a construct of desire which is wracked by expenditure, Cather states "the life-force seemed low" in the Pecos households (*DCA*, 122). There is a suggestion, as well, that the lack of desire in the pueblo has its source in a celibate

priesthood; the maintenance of a "ceremonial fire," apparently to source of a religious *élan*, and which "had from time immemorial" been kept in "some cave in the mountain behind the pueblo," paradoxically draws off, and has "sapped the strength of the young men appointed to serve it,--- always the best of the tribe" (*DCA*, 122). The image the reader has, then, of the Pecos household, suggests that it stands distinctly in opposition to the household of Padre Martinez, and indeed, one finds in the house of Jacinto a domestic space uncluttered by the reproduction of desire, if the implications of such a space remain troubling. There is but one woman in Jacinto's household, Clara, whose cooking is itself buoyed by tradition (she serves Latour "a basin full of hot corn-bread baked with squash seeds,--- an Indian delicacy comparable to raisin bread among the whites), and whose service is spotless and marked by pious composure (*DCA*, 121). There is also one child, of whom, when asked by Latour, Jacinto makes the sad reply "that the baby was ailing" (*DCA*, 121). Yet the space of the home is not altogether unlike the Episcopal residence, and here too, "burning pinion logs filled the room with sweet-smelling smoke," a room which, like the Bishop's study, seems sanctified in the "meanest tasks," and whose walls are "smoothly whitewashed, and clean, to the eye, at least, because of [their] very bareness" (*DCA*, 121). The room itself is furnished sparsely, an interiorized, and domesticated image of Manuel Lujon's *portales*, with "nothing on the walls but a few fox pelts and strings of gourds and red peppers" (*DCA*, 121). In place of the Bishop's prized books and artifacts, are "richly coloured blankets of which Jacinto was very proud,... folded in piles on the earth settle" (*DCA*, 121). As disconcerting as the fate of Pecos pueblo appears, with its "long rock ridges of dead pueblo,-- empty houses ruined by weather" slowly

encroaching upon the "living pueblo" adjoining, it remains the space of composure and reflection, untroubled by desire, attached to history, and sanctified by domestic order.

As drawn together by the images of domestic order as are the Native American, and the European residence, the two representative individuals, the Bishop, and his guide, seem drawn together by a mutual grace and refinement, altogether distant from the excessive animation of the Hispanic body as Cather has given it to us. The Bishop, thinking "with satisfaction that he was beginning to have some sort of human companionship" with Jacinto, admires the fact that the guide "was as much at home in the Bishop's study as in his own pueblo," and hardly the slow, sensual student of theology that was Trinidad, feels that "his training, whatever it had been, had prepared him to meet any situation which might confront him" (*DCA*, 93). In turn, Cather notes, "Jacinto liked the bishop's way of meeting people; thought he had the right tone," with his priests, and "that he had good manners with the Indians" (*DCA*, 93). Yet perhaps what is most refined in their relationship toward on another is the fact that "silence... was their usual form of intercourse" (*DCA*, 91). The sense of ease with which the two individuals spend time with one another is occasionally marked by brief conversation, yet for the most part, discourse tends to create the space of contemplation and reflection which comes to define the relationship. Reflecting *on* one another, reflected *in* one another, Cather seems to suggest that the Native American, and the European consciousness meet in an arena of receptivity which is never over-animated, but which, in its silence, is a truly contemplative form of cultural transmission.

However, Cather suggests as well that those moments of discourse which find their way back to the repose of contemplation are not always, in and of themselves, shaped in the midst of a pure receptivity and exchange of cultural world-views from one individual to another. To be sure, the exchange of silence, which so strongly binds the Native American to the European, is often occasioned by the fact that discourse can lead to the awareness of an untranslatable gulf in world-views, which is not so much a breakdown of communication as it is the creation of an Otherness in opposing individuals, which leaves each the "object" of reflection for the other. Making their camp on a rock ledge outside the pueblo of Laguna, both Jacinto and the Bishop seem to have fixed their attention "[h]igh above the horizon, "where the evening star flickered like a lamp just lit," and where, Cather notes, "close beside it was another star of constant light, much smaller" (*DCA*, 92). Cather draws attention to the possibility that the mutual gaze might be the occasion for some sort of cultural translation, but the guide seems unsure of his language. Jacinto, who speaks "without being addressed," and without gesture, apparently, utters "'[t]he ev-en-ing-star'... slowly and somewhat sententiously in English," and then somehow dissatisfied, Cather notes that "he relapsed into Spanish" to offer a further comment on "the little star beside" (*DCA*, 92). "Indians call him the guide," he tells the Bishop," and there is a clear sense in which the shifting ground of discourse which moves from English, to Spanish, and ultimately, toward an untranslatable Native American language evoked purely as a cultural symbol, constructs the space of contemplation whose focus is the relation of self and other (*DCA*, 92). Ultimately, the night sky reflects the image of priest and guide, as they make their camp on the rock ledge, where the "two companions sat,"

Cather notes, "each thinking his own thoughts as night closed in about them" (*DCA*, 92). The Bishop, who "seldom questioned Jacinto about his thoughts or beliefs," fully, and deeply ruminates on the untranslatability of world views, and in the silence, reflects:

> There was no way in which he could transfer his own memories of European civilization into the Indian mind, and he was quite willing to believe that behind Jacinto there was a long tradition, a story of experience, which no language could translate to him. A chill came with the darkness. (*DCA*, 92).

There is evoked here, a mild sense of dislocation in which the night sky as an image of cosmos, or of home, becomes a contested space in the assertion of two distinct cultural world-views. Jacinto, who seems quite literally to be "guiding" the Bishop through this meditation, breaks the silence, and expands the issue to encompass the breadth of the evening canopy. "Many stars," he suggests, and then asks the Bishop "What do you think about the stars, Padre?" (*DCA*, 92). Latour's response appeals to science rather than religion, and perhaps the Bishop despairs altogether of making the Christian narrative accessible to the Pecos guide. "The wise men tell us they are worlds, like ours," he states, though Jacinto replies abruptly "I think not" (*DCA*, 92). The Native American is much more comfortable espousing his faith in a metaphysical cosmos, and he tells the Bishop "I think they are leaders--- great spirits" (*DCA*, 93). The gulf which is opened by Jacinto's rejection of the Bishop's explanation both separates, and draws together the two individuals, insofar as it somehow gives the lie to Latour's rationalism as the essence of his own world-view, and re-asserts the cosmos as the site of a mystery, which, if understood in the context of different cultural world-views, remains the space of a shared vantage point. "Perhaps they are," the Bishop states

"with a sigh . Whatever they are, they are great" (*DCA*, 93). Before so vast a realization, it seems, there is little left to do but settle into that posture of repose which turns towards forgetfulness, and with a final assertion of that most economized narrative of Christian redemption, he tells Jacinto "[l]et us say *Our Father*, and go to sleep, my boy" (*DCA*, 93).

The repose of sleep, while fraught with the loss of conscious thought, and driven, perhaps, by a gesture of yielding before an incomprehensible mystery, remains an act of faith in the New Mexico desert, expressing as it does the desire that the vault of night sky enclosing the body at rest fashion a space which itself is a construct of home, of shelter and sanctuary. Latour and Jacinto create a wilderness domestic on the rock ledge in front of the pueblo, from which, Cather notes, "the smell of piñon smoke came softly through the still air," a sort of blessing of the transitory hearth which recreates the scene of the home, and makes of it the center of the world, so to speak (*DCA*, 91). Having decided against sleeping in the sacristy of the church at Laguna, a structure which is itself a contested ground of cultural hegemony, "painted above and about the altar with gods of wind and rain and thunder, sun and moon," the Bishop feels comfortable in the open air of the desert (*DCA*, 89). There is, in this transferal, and re-inscription of a naturalized art limning the space of the sanctuary, from interior, to exterior, a broadening of the sense of enclosure; in other words, if the Church at Laguna represents the attempt to enclose natural space about the images of Christian redemption, the Bishop's rejection of the sacristy as a place of repose fundamentally relocates this artistic discourse, anchoring it squarely in the reality of the world itself. Either space remains the ground of contested cultural views, and yet the artifice of

centering the altar in the middle of the world is naturalized in the image of the blessed hearth, and the cosmos itself takes on the image of sanctuary. More deeply inscribed in this gesture, however, is something of a nostalgia for what Georg Lukács describes in *The Theory of the Novel* as the "integrated civilization," an epic past marked with the wholeness of cultural tradition, understanding of the universe, and the abiding sense of being "at home" wherever one is. It is worth noting the opening lines of Lukács' chapter on epic literature for their own evocation of the "cosmological home" of a distant past, where he notes:

> Happy are those ages when the starry sky is the map of all possible paths--- ages whose paths are illuminated by the light of the stars. Everything in such ages is new and yet familiar, full of adventure and yet their own. The world is wide and yet it is like a home, for the fire that burns in the soul is of the same essential nature as the stars; the world and the self, the light and the fire, are sharply distinct, yet they never become permanent strangers to one another... [21]

Lukács, in a preface to the original draft, suggested that *The Theory of the Novel* was burdened throughout with this utopic impulse, but his attribution of the "mood" of the text to the dislocation experienced by the First World War, is interesting with respect to Cather's assertion that for her, "the world broke in two in 1922 or thereabouts."[22] Cather had completed her own novel on the war and its discontents with *One of Ours*, but though she suggested that *Death Comes for the Archbishop* was "like a happy vacation from life," asserting that the "happy mood in which I began it never paled," she seems to have continued, at moments in the text, to meditate on the figure of homelessness in search of a lost golden age of the "integrated civilization."[23] Wracked by spiritual despair after his Mass at the pueblo of Ácoma, Cather states the Bishop found himself:

> on a naked rock in the desert, in the stone age, prey to homesickness for his own kind, his own epoch, for European man and his glorious history of desire and dreams. (*DCA*, 103).

The moment is something of an idyll for a transatlantic past felt as the trauma of dislocation in the New World Missionary. Its source, however, lay in the dislocation of the historical consciousness confronted by a human prehistory which is at once preternatural, yet fraught with what the Bishop observes is a "strange literalness, often shocking and disconcerting" (*DCA*, 97). The "rock turtles" of Ácoma, before whom the Bishop "felt as if he were celebrating Mass at the bottom of the sea... for types of life so hardened, so shut within their shells that the sacrifice on Calvary could hardly reach back so far," are the embodiment of a physical necessity which motivates heroic human striving, yet which fails to attach itself to any redemptive schema other than the radical preservation of the life process in a hostile environment. "A man can do a whole lot when they hunt him day and night like an animal," Jacinto tells Latour when he wonders "how... did men first think of living on top of naked rocks like these, hundreds of feet in the air?" (*DCA*, 96). And yet celebrating mass, the Bishop reflects that those "shell-like backs behind him might be saved by baptism or divine grace... but hardly through any experience of their own" (*DCA*, 100). The nature of human strife on the rock of Ácoma is monumental, and yet belongs to prehistory; the experience of Christian redemption is of less worth here than the experience of protection, enclosure, and sanctuary which seems a part of the body itself, part of the "shell-like" repose which comes from isolating oneself in the midst of violence, rather than in the midst of divine grace.

As the Bishop and Jacinto approach Ácoma pueblo on horseback, their discussion leads Latour to the understanding that "[a]ll this plain had once been a scene of a periodic manhunt" (*DCA*, 97). Human history mediates this scene of mesa and bluff, and seems inscribed onto the face of the landscape itself. The Bishop, prone to the civilizing gaze, has been riding through a countryside in which the natural world resembles a vast *civitas*; his view "[f]rom the flat red sea of sand," Cather notes, is set toward:

> great rock mesas, generally Gothic in outline, resembling vast cathedrals. They were not crowded together in disorder, but placed in wide spaces, long vistas between. This plain might have once been an enormous city, all the smaller quarters destroyed by time, only the public buildings left---piles of architecture that were like mountains. (*DCA*, 94).

And yet the civilizing impulse is so clearly traumatized by the effects of time on the "small quarters" of this desert scene; with only the "public buildings left," the natural cityscape seems a monument to the *vita activa*, the space of the *polis* which, as Hannah Arendt notes, "produces beautiful deeds," but which, also, privatizes the sphere of the home, of domestic economy viewed as the problematic space where necessity makes itself felt in the maintenance of the life processes.[24] Birth and death, daily life, these are evoked as absence in the ruination of "small quarters," and what remains are the monuments to human ambition, public deeds, and ultimately, in the Gothic landscape, the "civic" nature of the Christian world-view, and its active "mission" in the New World. What is disconcerting in the scene, perhaps, is the very lack of the space of contemplation and reflection which has been made of the household in *Death Comes for the Archbishop*; the Bishop, who would come to feel as much the agent of a civilizing impulse driven by restlessness, by the

Church's need to hunt for lost souls, by the policy of American expansion, indeed wonders before this prehistorical scene of "a periodic man-hunt" born out of fear and violence, and yet so "civil" in its natural state.[25]

However, the Bishop manages to draw a deep correlation between the drama of "prehistory," and the drama of Christian redemption. Meditating on the idea that "these Indians, born in fear and dying by violence, had at last taken this leap away from earth," he feels assurance in the fact that "on that rock [they] had found the hope of all suffering and tormented creatures--- safety" (*DCA*, 97). There is a sort of ambition about this gesture, so amazing to Latour, and yet so fraught with worldly care and concern for the most fundamental needs of the life process itself. In its essentialist nature, the rock of Ácoma suddenly becomes the cornerstone for all human strife, longing, and desire, and the Bishop observes:

> The rock, when one came to think of it, was the utmost expression of human need; even mere feeling yearned for it; it was the highest comparison of loyalty in love and friendship. Christ himself had used that comparison for the disciple to whom he gave the keys of His Church. And the Hebrews of the Old Testament, always being carried captive into foreign lands, --- their rock was an idea of God, the only thing their conquerors could not take from them. (*DCA*, 97).

It's difficult to see an arrogance in the gaze of the Bishop as he looks over the scene before him; and yet there is something so startling about the manner in which the "periodic manhunt" of the New Mexico desert comes to valorize the incessant man hunt of the Judeo-Christian drama, as to appropriate the very physicality of pueblo life itself, to spiritualize the labor inherent in the realities of life on a rock, a reality which begs the question of what need these individuals have to remain up there, having

taken their leap in what ought to have been the prehistory, but hardly the present, of providential history in the New World.[26]

So it is that the worldliness of "prehistory", the incessant concern for the necessities of life, makes of this ambitious leap of the Ácomas into the Christian drama the hopeful, yet problematic space of redemption. Though he attempts to conceptualize, Latour understands too, that the "public," and monumental cityscape he sees in the desert constitutes as much the icon of monumental human labor and daily strife, as it does the image of great acts and deeds whose motives are born of the desire to endure somehow, after life itself gives out. "There was an element of exaggeration in anything so simple," he thinks, in the fact that:

> The Ácomas, who must share the universal human yearning for something permanent, enduring, without shadow of change,--- they had their idea in substance. They actually lived upon their Rock; were born upon it and died upon it. (*DCA*, 98).

The idea, in substance, is inherently the idea of shelter, sanctuary, and sustenance whose elemental nature is tied to the physicality of life. That this constitutes a "worldly ambition" in the text is, problematically, the means to appropriate it, utilize it somehow by making it a part of a "higher," and universal need. Sharing human yearning for something higher, in the eyes of the Catholic priesthood, the Ácoma's monumental labor, the worldliness of their concept of sanctuary, these can be freely drawn upon and appropriated for the physical manifestation of theChurch in the New World. The "old warlike church of Ácoma," the garden of Fray Baltazar (both monuments to labor appropriated for the spiritual man hunt of the Catholic Church), ultimately define the landscape at the top of the rock mesa, and ultimately, they too are fraught with the "worldly

ambition" of an Hispanic priesthood whose role in the text is always problematic.

The Church of Ácoma, "[g]aunt, grim, grey, and half-ruined,....more like a fortress than a place of worship," is an intense site of labor as the Bishop and Jacinto look it over (*DCA*, 100). Latour, who reflects that the Hispanic priesthood "was not altogether innocent of worldly ambition," wonders that these "Spanish Fathers' managed to "draft Indian labour for this great work without military support" (*DCA*, 101). The fact that the church is half-ruined creates something of an indecent exposure of the extent of labour gone into the making of the structure; here, no "deft palms" of an idealized, benign Native American craft create the space of reflection, but rather, the crude mass of unfinished wall and ceiling becomes the occasion for an awakening to the human effort extracted by building on top of a mesa. "Every stone in that structure," the Bishop understands, every handful of earth in those many thousand pounds of adobe, was carried up the trail on the backs of men and boys and women" (*DCA*, 101). Visible, too, are "the great carved beams of the roof," and Cather notes "Father Latour looked at them with amazement" (*DCA*, 101). Jacinto only shrugs and says they must have been carried from the San Mateo range on the backs of laborers, some fifty miles, leaving the Bishop to wonder what "[p]owerful men they must have been, those Spanish Fathers, to draft Indian labour without military support" (*DCA*, 101). Altogether, the church, the "large, thick-walled" cloister which "must have required an enormous portage from the plain," the garden created and watered for the Epicurean tastes of Fray Baltazar--- the entire top of Ácoma mesa is the sight of monumental labor appropriated for what seems indeed a set of worldly ambitions (*DCA*, 102). The French

missionary decides to spend his night in the "loggia--- roofed but with open sides;" perhaps there is an act of contrition in rejecting the "four feet of solid, windowless adobe" of the church (*DCA*, 102). However, the history of the Church "proper" in New Mexico remains highly problematic with respect to the past of the Hispanic priesthood, who, the Bishop comes to feel in such surroundings, "might well have forgotten the poor Ácomas... and believed themselves in some cloister hung on a spur of the Pyrenees" (*DCA*, 102).

And yet, it remains hard to overlook this image of worldly ambition understood as a projection of the European landscape hung upon a "spur" of New Mexico desert, since it so strongly evokes the border between Spanish and French culture. After all, it could hardly be said that the Jesuit presence in the great diocese left untouched the landscape of the desert Southwest; ambitious undertakings account for a marked remapping of Santa Fé itself, in *Death Comes for the Archbishop*, as well as much of the cultural topography of the entire region, and into the Rocky Mountains. There is, however, a clear sense in which Cather's attention to the French project for the physical manifestation of the Church draws upon both the perogative of building at the core of Catholic authority in the diocese, as well as utilizing, as a backdrop to ambitious works, the much more serene and picturesque landscape which surrounds the village of Santa Fé. The shift from a sublime, "warlike" architecture of the Spanish Fathers to the "old churches of the Midi," which Latour understands to be "the most beautiful in France," is essentially a finalization of cultural hegemony, understood as a triumph of architecture in the New World. The Cathedral of Santa Fé, that worldly ambition of the Bishop, who wishes "to leave nothing to chance, or to the mercy of

American builders," remains a question of taste which draws upon tradition no less than does Vaillant's onion soup (*DCA*, 240). To be sure, Vaillant reminds himself as Latour speaks of his plans for the church, that the Bishop's "ancestors helped to build Clermont Cathedral;... two building Bishops de la Tour back in the thirteenth century. Time," he tells the Bishop, "brings things to pass, certainly" (*DCA*, 242). One hardly suspects Vaillant to "deprecate the individual talents" of his Bishop, though, as Cather notes, whether the Cathedral was to be "Midi-Romanesque or Ohio German in style seemed to him of little consequence" (*DCA*, 243).

However, good taste in matters of church-building seems so strongly to call into question the transformation of landscape as an act of the individual will, that the gesture itself appears to make a Frenchman of the missionary, rather than a missionary of the Frenchman; in other words, the physical structure of the church is so tenuously linked in Vaillant's mind with the redemption of souls in the Southwest, that a question of its features can only make itself felt *as* a matter of personal, and national taste. "Our own Midi-Romanesque is the right style for this country," Latour tells, Vaiilant, and he goes further by suggesting that "[i]t would be a shame to any man coming from a Seminary that is one of the architectural treasures of France, to make another ugly church on this continent" (*DCA*, 240, 242). And yet Vaillant worries "[i]f you once begin thinking about architects and styles, Jean!," and further, implies that the imposition of the Midi-Romanesque on the New Mexico landscape is a matter of taste which remains bound to the urgencies of life moreso than Latour would care to admit (*DCA*, 240). "As for me," he tells his Bishop, "I see only what is under my nose.... I had no idea you were going in for fine

building, when everything about us is so poor--- and we ourselves are so poor" (*DCA*, 241). There are two radically disparate perspectives on the landscape being contested here; the New Mexico desert is both beautiful and impoverished--- serene, and yet wracked by scarcity--- redeemed by the calm , far reaching vision of aesthetic gaze, and yet waiting to be redeemed by the labor of the missionary. "You plan far ahead," Vaillant tells Latour, and he admits "that is what a bishop should be able to do" (*DCA*, 241). And to be sure, Latour is able to offer assurance that "the Cathedral is not for us, Father Joseph. We build for the future--- better not lay a stone unless we can do that" (*DCA*, 240-41). Yet there is little doubt that, while "fine building" for the future is the construct of an aesthetic gaze which sees redemption as a *certain* prospect, or vista of landscape made humane by art, its will to power remains a matter of self-indulgence. Speaking of the rock with which he plans to build, Latour states:

> I could hardly have hoped that God would gratify my personal taste, my vanity, if you will, in this way. I tell you, *Blanchet*, I would rather have found that hill of yellow rock than have come into a fortune to spend in charity. The Cathedral is near my heart, for many reasons. I hope you do not think me very worldly. (*DCA*, 243)

Latour tells Vaillant that he has hopes of completing the cathedral before his death, and indeed, he had from the beginnings of his desires to build, Cather notes, a feeling that "such a building might be a continuation of himself and his purpose, a physical body full of his aspirations after he had passed from the scene" (*DCA*, 175). There is little doubt that Latour pursues the project of the cathedral as a sort of *homo faber*, whose work of the hands is bound up with the production of material things in the world which can survive the ephemeral nature of

life, even as it shelters, secures, and facilitates its activities during its time on earth. It is interesting to note that Cather tends toward an elision of the physical labor inherent in building such a structure; indeed, the concept of the "builder" tends to become lost in a hierarchy of authority which leads from Latour as the arbiter of taste in matters of design, to Molny, the French architect, to French stone-cutters, and finally, to an abstraction of physical labor which is only introduced in the text as a matter of speculation. "This hill is only about fifteen miles from Santa Fé," Latour tells his friend, "there is an upgrade, but it is gradual. Hauling the stone will be easier than I could have hoped for" (*DCA*, 241). Clearly, there is an evocation of the massive labor required in hauling materials up to the top of Ácoma mesa, and Cather does seem to suggest that there is an equivocal drafting of labor with regard to the Cathedral at Santa Fé. And yet, the completion of the structure so effectively blesses the landscape--- a structure which, Cather notes, "seemed to start so directly out of those rose-coloured hills--- with a purpose so strong that it was like action," that its role in the redemptive project for New Mexico Territory seems to blend the aesthetic with the religious in a benign image of the cultural transfiguration of an entire landscape. Both redeemed, and beautiful, the city of Santa Fé takes upon itself the aura of the *Civitas Deis* in the New World, and Latour as momentary *homo faber* is essentially the cultural, and religious architect of redemption in a desert landscape. That he will find in his own cathedral the space of his ultimate repose is no small matter; housed by the work become artifact, enclosed by landscape, by stone, and finally become "his own story," so to speak, the repose of the body reflects on a life of purpose, of action, of the work of redemption (*DCA*, 182).

[1] Willa Cather, *Death Comes for the Archbishop*. (New York: Vintage Books, 1990), 35. Hereafter, this text will be cited as *DCA*.

[2] The discussion of the Bishop's study on pages 33-34 of the text, is clearly an intent to structure an aesthetics of traditional New Mexican architecture; and one which does so by valorizing, and "beautifying" the marks of the hand made apparent on physical surfaces, and "shaped" spaces. That the study houses the Bishop's library, and his most prized religious artifacts is clear as well. The issue, however, lies again with the way in which the products of human labor interact with the symbols of the human mind, and spirit. Along with the theological books resting on the ax-hewn planks in the study, Cather notes that "[o]n either side of the fireplace, plastered recesses were let into the wall. In one, narrow and arched, stood the Bishop's crucifix. The other was square, with a carved wooden door, like a grill, and within it lay a few rare and beautiful books. The rest of the Bishop's library was on open shelves at one end of the room" (*DCA*, 34). While the image begs the question of whether the books, and the crucifix lend grace and beauty to the spaces made by human labor, or whether the opposite is true, it remains the case that the Bishop's possessions are so securely housed, sheltered, and supported by works of the hand, that an ontological "dilemma," of sorts, is opened early in the text, one which problematizes the very notion of the duality of mind and body so central to Western thought. A general discussion of this duality in American thought, and its relevance to the division of labor into manual and intellectual activity, along with the subsequent valorization of the primacy of intellectual labor, can be found in Nicholas K. Bromell's *By the Sweat of the Brow: Literature and Labor in Antebellum America*. (Chicago: University of Chicago Press, 1993), esp. pp. 15-60.

[3] Cather's unpublished fragment, "Light on Adobe Walls," is something of an enigma, insofar as it never explicitly mentions adobe as an aesthetic/reflective surface, and it is difficult to speculate what intentions she had with regard to this piece. She does note, however, with respect to the artistic representation of light (sunlight, in this instance) that "[n]obody can paint [it]. He can only paint the tricks that shadows play with it, or what it does to forms. He cannot even paint those relations of light and shade--- he can only paint some emotion they give him, some man-made arrangement of them that happens to give him personal delight." In *Willa Cather on Writing: Critical Studies on Writing as an Art*. (Lincoln, NE: University of Nebraska Press, 1988), pp. 123-26, cited, pp. 123-34. The woman artisan clearly arranges the surface of the adobe walls, even as Cather arranges the representation of light in a pleasing image that gives itself over directly to the Bishop in his study. At any rate, *artist* and *artisan* are both drawn together here in the construct of an emotive space. See also Judith Fryer, *Felicitous Space: The Imaginative Structures of Edith Wharton and Willa Cather*. (Chapel Hill: The University of North Carolina Press, 1986), esp. pp. 289-318 for a good discussion of Cather's use of Southwestern architecture and landscape as emotive spaces. Fryer's discussion lends itself to an idealized primitivism with respect to the desert southwest, though in some sense she is responding to those elements of Cather's work that are in themselves primitivist. I do not think that the scene of "The Bishop Chez Lui" attempts to draw so fully on the adobe space as the site of a "connectedness" to "traditional cultures," though that is certainly the case elsewhere in the text. Rather, I mean to suggest that the idea of "Felicitous Space" as "the space that concentrates being within limits that protect," is here the source of composure, or repose which has its source in the mediation of physical labor, and the aesthetic impulse which responds "warmly" to the play of light upon the hand-made surface. Fryer, p. 310.

[4] I mean to imply here, that Western art is, in my opinion, motivated in this text by an economy of expenditure inherent in an ideology of "Art for Art's Sake," one which suggests that effort expended in the production of aesthetic values serves no other purpose than the communication of the human spirit, and its desire to elevate itself above the experiences of a life dictated by conditions of scarcity and subsistence. It is

hardly my intention to argue this matter here; the assumption that art appears in those moments when a culture finds enough leisure in its conditions of existence to spend time in the production of aesthetic products is familiar enough, and suspect enough, I would add as well. That Cather occasionally indulged in this view, however, is evident, and that she further understood it to be a sort of transcultural component in the emergence of human civilizations, is clear also. This view, with respect to Southwestern indigenous cultures is perhaps the most clearly stated in *The Song of the Lark*, where Thea Kronberg, Cather's emergent artist, spends her days in Panther Canyon, near Flagstaff, Arizona, examining the ruins and artifacts of Anasazi culture. Examining the potsherds left behind by Native American women, Thea experiences a pervasive sense of what she perceives to be an originary moment in artistic endeavor. Cather states that "[n]early every afternoon [Thea] went to the chambers which contained the most interesting fragments of pottery, sat and looked at them for awhile. Some of them were beautifully decorated. This care, expended upon vessels that could not hold food or water any better for the additional labor put upon them, made her heart go out to those ancient potters. They had not only expressed their desire, but they had expressed it as beautifully as they could. Food, fire, water, and something else--- even here, in this crack in the world, so far back in the night of the past! Down here at the beginning, that painful thing was already stirring; the seed of sorrow, and so much delight." Willa Cather, *The Song of the Lark*. (New York: Signet Classics, 1991), pp. 263-264. Hereafter cited as *SOL*. That Thea sees her own endeavors as a culmination of this early expenditure is clear as well. "All these things," she reflects, "made one feel that one ought to do one's best, and help to fulfill some desire of the dust that slept there.... In their own way, those people had felt the beginnings of what was to come. These potsherds were like fetters that bound one to a long chain of human endeavor (*SOL*, 264). While true to the ideology of the redemption of human effort through art, and an invaluable statement with regard to Cather's interests in the female artist/artisan, I remain troubled by what seems to me a clear colonization of Native American women's labor as a means to fashion some sort of incipient moment in the growth of the Subject of modernity. For my part, I have little trouble considering the making of pottery, and the expenditure of time in its design as a fundamentally social activity, hardly born of sorrow, or desire; or that art, in some sense may have been born of distraction, or a fascination with a medium and a penchant for experimentation. At any rate, this aspect of Cather's fascination with Southwestern art and culture needs to be dealt with, and I raise the issue here primarily as a means to establish a cautionary view with respect to Cather's depiction of Anglo-European characters and their outlook on the Desert Southwest and its cultures.

[5] John Murphy is one critic who notes this duality inherent in the Bishop's appearance in the text's initial Southwestern scene. However, he argues that the "buckskin" apparel which hides the Bishop's religious garments place Latour in the tradition of Natty Bumppo or Owen Wister's Virginian, an image which leans too strongly toward the idealization of the Western Hero. Murphy suggests that "Wister's tall stranger and Cather's young priest are knights-errant in the wilderness, rescuing those in need, righting wrongs, bringing law to lawless regions." In my view, the image of the American is far too equivocal in the text to suggest that the ideal hero, and the ideal priest blend so comfortably in the figure of Latour. I'm prepared to take the Bishop at his word when he says he feels more like a Yankee trader than a model frontiersman, and less prepared to suggest that Cather felt too comfortable clothing her missionary in the standard dress of the American presence in Southwestern history. See John J. Murphy, "Willa Cather's Archbishop: A Western and Classical Perspective," in Harold Bloom, ed., *Modern Critical Views: Willa Cather*. (New York: Chelsea House Publishers, 1985), pp. 161-69, cited p. 162. See also, Josiah Gregg, *Commerce of the Prairies*. (New York: J.B. Lippincott Company, 1962) vs. 1&2, for an interesting "counter-history" with regards to the development of Santa Fé in the middle of the 19th Century. Gregg's perspective is essentially that of the Yankee-trader, again, a figure whose resemblance Latour sees in himself, and which he suggests is unrefined.

[6] Willa Cather, "On *Death Comes for the Archbishop*," in *Willa Cather on Writing*, pp. 3-13, cited p. 7. The definitive biography of Lamy's life is Paul Horgan's *Lamy of Santa Fé*. (New York: Farrar, Straus and Giroux, 1975).

[7] This investiture in domestic labor as a repository of cultural value, and potential site for the historical consciousness continues to disturb me for obvious reasons. And yet it would be a mistake to understate Cather's commitment to the home as a primary venue for the maintenance and transmission of the European world-view. This facet of Cather's work may be tempered elsewhere, but there is little mistaking the cultural impetus driving a novel such as *Shadows on the Rock*, Cather's second work meditating on French Catholic culture in the New World. In what is perhaps the most seductive imagery of domesticity I have seen in a modern novel, the space and artifacts of the kitchen are there the occasion for the realization by the young Cécille that experience is an idealized form of domestic labor, that "[t]hese coppers, big and little, these brooms and clouts and brushes, were tools and with them, one made, not shoes or cabinet-work, but life itself. One made a climate within a climate; one made the days... the special happiness of each day as it passed; one made life." Willa Cather, *Shadows on the Rock*. (New York: Vintage Books, 1971), p. 198. Here, too, we see Cather's strongest statements of domestic order as a primary civilizing influence, as the dying Madame Auclair urges Cécille to keep a good house in the New World. "[I]n time, she tells her daughter, "you will come to love your duties, as I do.... Without order our lives would be disgusting, like those of the poor savages. At home, in France, we have learned to do all these things in the best way... and that is why we are called the most civilized people in Europe and other nations envy us" pp. 24-25. See also Helen M. Dennis, "Good Housekeepers: Flaubertian Fictions of Domesticity," in Helen Dennis, ed. , *Willa Cather and European Cultural Influences*. (Lewiston, NY: The Edwin Mellen Press, 1996), 121-42, for a discussion of domesticity in *Shadows on the Rock*. The home is perhaps only the nexus for Cather's ideal of order in the New World; the landscape at large often seems to require a civilizing influence as well. But it remains the fact that the space of the home has more at stake with regard to cultural interaction than any other space in Cather's fiction, and its European outlook is fundamental regardless of her treatment of the domestic spaces and habits of other cultures.

[8] I mean to imply that the idea of Necessity inheres in the concept of "grace," either as a blessing of Providence in the Christian view, or idealized in the model of the "graceful," redemptive features of European high culture, as far as Cather is concerned. That a concept of grace can adhere itself to the physical body, is something I suggested in the last chapter; yet here, too, the individual who is able to overcome, through conceptualization and meditation, the physicality of need, is a graceful individual. Again, both "gentle-birth," and Christian devotion lend the features of grace to the Bishop lost in the desert, and as he prepares for his devotions, Cather notes that Latour "had a kind of courtesy toward himself, toward his beasts, toward the juniper before which he knelt, and the God whom he was addressing" (*DCA*, 19). Susan Rosowski has argued that this "scene dramatically illustrates the saving power of conceptualization," and that Latour's ability to construct the symbol of redemption by kneeling before the "cruciform" juniper gives meaning and order to the endlessly repetitive landscape. See Susan J. Rosowski, *The Voyage Perilous: Willa Cather's Romanticism*. (Lincoln, NE: University of Nebraska Press, 1986), p. 172. I tend to agree, but I would argue as well that this scene ,which ruminates so heavily on the refined gesture, and the physical image of piety, offers meditation as a redemption of the body as well, and erasure of the possible ruination of the individual wracked by hunger, or thirst. "His devotions lasted perhaps half an hour," Cather notes of the Bishop, "and when he rose he looked refreshed" (*DCA*, 19). The scene so effectively negates the destruction of the physical generally associated with the Catholic *imitatio Christi* in the form of scourging, the *stigmata*, and, notably, the "scandalous" rites of the Penitential Brotherhood in New Mexico during Passion week, that it seems for the moment, to come out on the other side of the sacrifice at Calvary, so to speak, beyond the cry of *J'ai Soif!*" wrung on the cross. See *DCA*, "The Old Order," pp. 139-58 for a discussion of the Penitentes, especially the behavior of Padre Martinez' "nephew " Trinidad, ,whose sloth, gluttony, and physical features are in direct opposition to those of one such as the Bishop. Señora Carson, in speaking of Trinidad, suggests anything but a graceful physical piety in his own relation

to the cross, and tells Latour that "at Abiquiu last year, in Passion Week," Trinidad "had himself crucified....He was tied upon a cross with ropes, to hang there all night.... But he is so heavy that after he had hung there a few hours, the cross fell over with him, and he was very much humiliated" (*DCA*, 154). See also Ann W. Fisher-Wirth, "Dispossession and Redemption in the Novels of Willa Cather." *Cather Studies*, vol. I, 1990, 36-54, for a further discussion of this scene as a moment in which the desert landscape is redeemed by Latour's symbolization of the cruciform juniper.

[9] Hannah Arendt's discussion of the relation between the *vita activa* and the *vita contemplativa* in both classical and Christian thought is useful in examining the tension between the need to be active, and the need for contemplation in the life of the missionary. See Hannah Arendt, *The Human Condition.* (Chicago: The University of Chicago Press, 1858), esp. pp. 14-17. Most cogently, perhaps, her brief note that "Augustine speaks of the 'burden' (*sarcina)* of active life imposed by the duty of charity, which would be unbearable without the 'sweetness' (*suavitas*) and the 'delight of truth' given in contemplation (*De civitate Dei xix. 19)*" speaks to Vaillant's poignant deferral of devotion to Catholic Mariology, which is associated elsewhere in *DCA* with the imagery of rest, and recuperation. Cited, n. 14, p. 16. See "The Month of Mary," (*DCA*, 199-210).

[10] Hermione Lee, *Willa Cather: A Life Saved Up.* (London: Virago Press, 1989), p. 264. Sally Peltier Harvey relies heavily on the ideal of the epic hero in her reading of *Death Comes for the Archbishop*, though she tends to subjectivize the concept, arguing that the text is "both a journey toward self-understanding and an epic adventure of religious devotion." Sally Peltier Harvey, *Redefining the American Dream: The Novels of Willa Cather.* (Rutherford, NJ: Fairleigh Dickinson Press, 1994), p. 95. I do not mean to imply that her reading of the text ignores the issues of diverse cultural perspectives as the work's "social," and even historical basis. Yet the central thematic concern remains a sense of self realization through community, which at times fails to take fully to task the text's cultural imperialism, and undeniable valorization of European culture. Rather, I would suggest that the quality of epic in *Death Comes for the Archbishop* lay more securely with Mikhail Bakhtin's argument that "national tradition (not personal experience and the free thought that grows out of it) serves as the source for the epic," as well as "its reliance on impersonal and sacrosanct tradition, on a commonly held evaluation and point of view." See "Epic and Novel," in M.M. Bakhtin, *The Dialogic Imagination*, ed. Michael Holquist. tr. Caryl Emerson and Michael Holquist. (Austin, TX: University of Texas Press, 1981), pp. 3-40, cited, p. 13, and p. 16. Cather herself was "amused" that the work was hard to classify, and in response to critics who "assert[ed] vehemently that it is not a novel," said "[m]yself, I prefer to call it a narrative." *Willa Cather on Writing*, p.12.

[11]Gillian Brown's "Getting in the Kitchen With Dinah: Domestic Politics and Uncle Tom's Cabin," is a landmark work examining the relation between domestic economy, and modes of production in society at large. Her argument that the chaos of Dinah's kitchen reflects the chaotic, and irrational forces at work in the institution of slavery is, of course, too strong a case with regard to the mode of production of *ranchero* culture in New Mexico. However, it is clear that the image of Manuel Lujon's ranch, "like a little town, with all its stables, corrals, and stake fences," while a clear representation of wealth and productivity, nevertheless exhibits a sort of carelessness about the house, under whose *portale* Cather notes the wall was "hung with bridles, saddles, great boots and spurs, guns and saddle blankets, strings of red peppers, fox skins, and the skins of two great rattlesnakes" (*DCA*, 54). The masculine nature of these items, except for the string of peppers, suggests that frontier habits blend too readily, perhaps, with the domestic space, and further, there are no images of piety in the household,; no household saints are mentioned, and the *sala* is hastily improvised as a space for Father Vaillant's religious devotions. See Gillian Brown, *Domestic Inidivuidualism: Imagining Self in 19th Century America.* (Berkely, University of Cailifornia Press, 1990).

[12] Evelyn Helmick Hively, *Sacred Fire: Willa Cather's Novel Cycle*. (Lanham, MD: 1994), p. 8.

[13] Cather's religious faith hardly seems to have been unswerving; however, it seems unlikely to me, with respect to her later novels, that she cared to envision a world totally given over to faith in human intelligence, art, and science. See Mildred R. Bennett, *The World of Willa Cather*. (Lincoln, NE: The University of Nebraska Press, 1951) for a discussion of Cather's conversion, as well as other aspects of her religious thoughts and views, esp. "Faith is a Gift," pp. 134-37.

[14] The Archbishop's late reflections on the Canyon de Chelly, and the role of his friend Kit Carson, is perhaps an exception here; more so perhaps because the canyon itself is so clearly a naturalized and feminized domestic space, which the Navajo chief Manuelito states is the place in which "his people had lived when they were a small weak tribe; it had nourished and protected them; it was their mother" (*DCA*, 292). Latour, essentially, remains powerless in the face of American policy; however, the restoration of the canyon, and its reflowering as an "Indian Garden of Eden" is so strong an image as to undo the events of the past, for all intents and purposes. The Archbishop's final trip to the canyon, mediated by the aesthetic gaze of the tourist as Latour and his French architect "behold the strange cliff dwellings," of tradition blending with the images of a healthy present, remains a sort of colonization of the domestic space. (*DCA*, 295). "Once more," he reflects, "crops were growing down at the bottom of the world between towering sandstone walls; sheep were grazing under the magnificent cottonwoods and drinking streams of sweet water" (*DCA*, 295). The pastoral is essentially a wilderness domestic for the Native American, Cather seems to imply, the more so since the towering sandstone walls are so strong an image of enclosure. See Elizabeth Ammons, *Conflicting Stories: American Women Writers at the Turn into the Twentieth Century (New York: Oxford University Press*, 1992), esp. "Art: Willa Cather, the Woman Writer as Artist, and Humishuma," pp. 121-39 for a discussion of Cather's problematic representations of race

[15] Willa Cather, "On *Death Comes for the Archbishop*, " in *Willa Cather on Writing*, p. 5.
[16] Anyone reading *Death Comes for the Archbishop*, should examine more objectively the historical figure of Padre Martinez. *Padre Martinez: New Perspectives from Taos*, is a good collection of essays, all of which ask the reader to reconsider Cather's treatment of this individual.

[17] Willa Cather, "On *Death Comes for the Archbishop*," in *Willa Cather on Writing, op. cit.*, p. 5.

[18] *ibid.*, p. 5.

[19] That Martinez' "schismatic" views can be so strongly placed in opposition to both American, and European culture and authority, is something Cather suggests on page 139, where she notes: Both the priest and people there were hostile to Americans, and jealous of interference. Any European, except a Spaniard, was regarded as a gringo" (*DCA*, 139). Drawing Latour into the mainstream of American culture, as a gringo, is certainly problematic, given the text's antipathy toward American culture and attitudes. Yet, he is somewhat fated to be an agent of expansion, and I raise the issue for the most part to "segregate" my use of the term Hispanic from any participation in what is meant in the text by a concept of the European, or of "western man." Again, I'm uncomfortable with this, but I see little point in arguing Cather's views with respect to her constructions of ethnicity.

[20] Guy Reynolds, "Willa Cather's Translated World: Migration and Linguistic Diversity," in Dennis, *Willa Cather and European Influneces*, pp. 97-110, cited p. 103.

[21] Georg Lukács, *The Theory of the Novel.* (Cambridge, MA: MIT Press, 1971), p. 29.

[22] See Lukács, *Theory of the Novel,* pp. 11-23. The quotation of Cather comes from the Prefatory note to *Not Under Forty.* (New York: Alfred A. Knopf, 1988).

[23] Willa Cather, "On *Death Comes for the Archbishop*," in *Willa Cather on Writing*, p. 11, and p. 10.

[24] See Arendt, *The Human Condition*, esp. "The Public and the Private Realm," pp. 22-78, for a discussion of the valorization of the public realm of deed and action, and the "hiddenness" of the biological imperatives of life within the domestic space, in the Greek conceptualization of the city-state.

[25] Father Vaillant, whose motive is "Rest in Action," is always the primary agent in the Catholic "man hunt" carried on by the New World missionary. See, "The Month of Mary," pp. 199-209 for a discussion of the desire for the life of contemplation, which seems troubled by a restlessness inherent in the missionary impulse. It is here, that Vaillant, devoting himself to contemplation of the Virgin, primarily because he is convalescent, expresses the urge "to hunt for lost Catholics.... Utterly lost Catholics" (*DCA*, 206). I see no reason not to extend the idea of the "utterly Lost Catholic" to the Native American, as far as the Catholic mission in the New World is concerned. For a discussion of Latour's equivocal response to the civilising process he himself has been a part of, see (*DCA*, 273), where he reflects "the peculiar quality in the air countires vanished after they were tamed by man and made to bear harvests.... The moisture of plowed land, the heaviness of labour and growth and grain-bearing, utterly destroyed it; one could breathe that only on the bright edges of the world, on the great grrass plains or the sage-brush desert."

[26] I know the term "prehistory" with respect to the periodic manhunt, and the pueblo life of the Southwest, is highly problematic; I use the term however, in what I suspect is the concept of any arena of life in this text, which operates outside the European, and Christan contexts of history.

CHAPTER 5

HAVING DONE WITH CALENDARED TIME: REPOSE, REDEMPTION, AND THE IDEOLOGY OF PROGRESS

The final image of *Death Comes for the Archbishop*, that of the final morning which finds the Archbishop of Santa Fé in state "before the high altar of the church he had built," is in many respects an image of the culmination of the individual's progress toward redemption.[1] Secure in the knowledge of Christian grace, yet secure too, within the stone walls of his European architectural triumph, the Archbishop in state is the consummate realization of the body in repose, beyond desire, and survived by the works of the hands which ensure his role in a broader, and more "collective" project of redemption. And yet, for the individual whose life has been inscribed with a plenitude of experiences— who, like the Archbishop, understands death not as a condition wrought by the ruination of the body, but rather, as the logical outcome "of having lived," the true measure of repose seems to be measured as a shifting ground of consciousness that precedes death itself as an experience of the ego which finds itself, altogether out of time, and yet replete with the understanding of one's own life as a story, a progressive narrative in its own right. (*DCA*, 267). Before the Archbishop "loses" his life, he finds it, or rather, Cather suggests:

> he sat in the middle of his own consciousness; none of his former states of mind was lost, or outgrown. They were all within reach of his hand, and all comprehensible. (*DCA*, 288).

The plenitude of experience, the psyche as the central organizing principle of one's own life story, of one's own "works" in the world--- this, in reality constitutes the true measure of grace as a state of conscious repose.

And yet, insofar as the Archbishop finds himself "out of time," or rather in the middle of it, within reach of all experiences from a perspective of the *nunc stans* imaged as the "end" of one individual's progress toward redemption, so too, perhaps, has the redemptive narrative of Catholicism, and to a great extent, the whole cultural experience of the American Southwest, found itself too easily understood to be outside, beyond, or behind the process of history as it has been written in the larger vistas of Protestant America. The erasure of desire, though rife with the imagery of the pastoral, the idyllic, the timelessness of plenitude, has hardly come to be the defining image of the American landscape. Donald Worster has suggested, rather, that, where the West has made itself felt as the landscape of scarcity, it has become, paradoxically:

> an appealing prospect in the white American, or in the white European mind. We have re-organized our whole way of life, our institutions and laws, to triumph over it. I maintain, in fact, that the drive to overcome the fear of scarcity has been one of the greatest forces pushing us toward the modern world.[2]

The idealization of progress, then, seems to ultimately have become intimately tied into the forms of desire; the fear of scarcity, in other words, tends more towards the western landscape's uncanny insistence on recycling the images of scarcity even as more and more, it becomes the "heavenly" dwelling place of a larger and larger share of the population. The history of redemptive narratives in the American west has been a history of seduction, to be sure--- the ideology of the body at rest, of natural

plenitude, strong markets, economic booms, all these images have found their way into narratives of western progress. However, it is the "ends" of progress that truly come back to haunt us, those images of plenitude which tend toward stasis, toward the ennui of life in eternity.

---

[1] Willa Cather, *Death Comes for the Archbishop*. New York: Vintage Classics, 1990), p. 297. Hereafter cited as *DCA*.
[2] Donald Worster, *Under Western Skies: Nature and History in the American West*. (New York: Oxford University Press, 1992.

BIBLIOGRAPHY

Primary Sources

Books:

Anello, Marc, comp. *Gone Fishin'*. White Plains, NY: Peter Pauper Press, 1996.

Corthell, Mrs. N.E., and Hill, Mrs. John A. *A Family Trek to the Yellowstone, and Twenty-four Years After*. Laramie, WY: Laramie Printing Co., 1923. Yellowstone National Park Library.

Banks, Nancy Houston, and Northrop, Henry Davenport. *The World's Fair as Seen in One Hundred Days*. Cincinnati, OH: The Standard Publishing Co., 1893.

Cather, Willa. *Death Comes for the Archbishop*. New York: Vintage Books, 1990.
---. *Not Under Forty*. New York: Alfred A. Knopf, 1988.
---. *Shadows on the Rock*. New York: Vintage Books, 1971.
---. *The Song of the Lark*. New York: Signet Classics, 1991.
---. *Willa Cather on Writing: Critical Studies on Writing as an Art*. Lincoln, NE: University of Nebraska Press, 1988.

Chapple, Joe Mitchell. *A' Top O' the World: Wonders of the Yellowstone Dreamland*. Boston: Chapple Publishing Co., 1922. Yellowstone National Park Libnry.

Chittendon, Hiram Martin. *The Yellowstone National Park*. Cincinnati, OH: The Robert Clark Company, 1909.

Cooper, James Fenimore. *The Pioneers*. New York: Bantam Books, 1993.

Cotton, Charles, and Walton, Isaac. *The Compleat Angler*. New York: Oxford University Press, 1982.

Crèvecoeur, Hector St. John de. *Letters From an American Farmer*. New York: E.P. Dutton and Co. Inc., 1945.

Davis, Rebecca Harding. "Life in the Iron Mills." in Ed. Nina Baym et. al. *The Norton Anthology of American Literature*. Shorter Edition. New York: W.W. Norton & Co., 1989: 1123-49.

Farnham, Eliza W.. *Life in Prairie Land*. Urbana, IL: University of Illinois Press, 1988.

France, Lewis B. *With Rod and Line in Colorado Waters*. Denver, CO: Chain, Hardy & Co., 1884.

Gregg, Josiah. *Commerce of the Prairies*. 2 vols. New York: J.B. Lippincott Co., 1962.

Irving, Washington. *A Tour on the Prairies*, Ed. John Francis Mc Dermott. Norman, OK: University of Oklahoma Press, 1956.

Kipling, Rudyard. *From Sea to Sea, Letters of Travel*. 2 vols. New York: Charles Scribner's Sons, 1906.

Kirkland, Caroline. *Forest Life*. 2 vols. New York: C.S. Francis and Company, 1842. Rpt. by Literature House, Upper Saddle River, NJ, 1970.

---. *A New Home, Who'll Follow or, Glimpses of Western Life*. New Brunswick, NJ: Rutgers University Press, 1990.

Langford, Nathanial Pitt. *The Discovery of Yellowstone National Park*. Lincoln, NE: University of Nebraska Press, 1972.

Maclean, Norman. *A River Runs Through It, and Other Stories*. New York: Pocket Books, 1992.

Mohr, Nicholas. Ed. Ray Allen Billington. Trans. La Vern J. Ripley and Klaus Lanzinger. *Excursion Through America*. Chicago: R.R. Donnelley & Sons Co., 1973. Yellowstone National Park Library.

Norton, Harry J. *Wonder-Land Illustrated*. Virginia City, MT: Harry J. Norton, 1873. Yellowstone National Park Library.

Richards, Mary Bradshaw. *Camping Out in the Yellowstone*. Ed. William W. Slaughter. Salt Lake City, UT: University of Utah Press, 1994.

Russell, Osborne. Ed. Aubrey L. Haines. *Journal of a Trapper*. Lincoln, NE: University of Nebraska Press, 1955.

Schullery, Paul. Ed. *Old Yellowstone Days*. Boulder, CO: Colorado Associated University Press, 1979.

Stanley, Edwin J. *Rambles in Wonderland*. New York: D. Appleton & Co., 1880. Yellowstone National Park Library.

Strahorn, Carrie. *Fifteen Thousand Miles by Stage*. New York: G.P. Putnam's Sons, 1911.

Tocqueville, Alexis de. *Democracy in America*. 2 vols. Ed. Phillips Bradley. New York: Vintage Classics, 1990.

Warren, Henry White. *Among the Forces*. New York: Eaton & Mains, 1898.

Wheeler, Olin D. *Wonderland 1904*. St. Paul, MN: Northern Pacific Railway, 1904. Haynes Special Collections, Renne Library, Montana State University.

White, Alma. *With God in the Yellowstone*. Zarepath, NJ: Pillar of Fire, 1933.

Articles and Reviews:

Allen, Margaret Andrews. "A Family Camp in Yellowstone Park." *Outing* (November, 1885): 157-59.

Rev., "*A New Home, Who'll Follow or, Glimpses of Western Life*, by Mrs. Mary Clavers, an Actual Settler." *Knickerbocker Magazine* XIV (November, 1839): 452.

Carnifex, John. "Fishing Along the Pecos River." *Outing* (July, 1891): 1-8.

Everts, Truman C. "Thirty-seven Days of Peril." *Scribner's Monthly*. 3 (November, 1871): 1-17.

Gordon, Theodore. "A Little Talk About the Angler's Flies." *Forest and Stream*. LXXIX (20) (November 16, 1912): 626-27.

Hamp, Sidford. Ed. Herbert O. Brayer. "Exploring the Yellowstone with Hayden, 1872: Diary of Sidford Hamp." *Annals of Wyoming* vol. 14 no. 4 (October, 1942): 253-98.

Hayden, Ferdinand. V.. "The Hot Springs and Geysers of the Yellowstone and Firehole Rivers." *The American Journal of Scinece and Arts*. Third Series, vol. III, no. 13: 105-15.

"Opening of the Great Fair." *Harper's Weekly* XXXVII (May 13, 1983): 442.

"The Power Plant." *Harper's Weekly* XXXVII (May 13, 1893): 442.

Van Dyke, T.S. "Among the Golden Trout." *Forest and Stream* vol. XLVI no. 10 (March 7, 1896): 442-45.

Unpublished Materials:

Cruikshank, Margaret Andrews. Ed. Lee H. Whittlesey. "Mortal Language Seems Entirely Inadequate: A Lady's Trip to Yellowstone in 1883." Yellowstone National Park Library.

Hague, Arnold. Letter to R.B. Moore, December 14, 1908. National Archives, U.S. Geological Survey, Record Group 57, Hague Papers, Entry 68.

Hughes, Thomas McKenny and Mrs. Caroline. "Excerpts from the Hughes Diary While Attending an International Geological Congress in America in 1891." Yellowstone National Park Library.

Pardo, Dorothy Brown. "Dorothy in Wonderland." Typed Manuscript. n.d. Yellowstone National Park Library.

Sanders, Wilbur Edgarton. "Journal of a Trip to Yellowstone, August 19-September 8, 1881." Typed Manuscript. Yellowstone National Park Library.

Secondary Sources

Books:

Ammons, Elizabeth. *Conflicting Stories: American Women Writers at the Turn Into the Twentieth Century*. New York: Oxford University Press, 1992.

Appleby, Joyce. *Capitalism and the New Social Order: The Republican Vision of the 1790's*. New York: New York University Press, 1984.
—. *Liberalism and Republicanism in the Historical imagination*. Cambridge: Harvard University Press, 1992.

Arendt, Hannah. *The Human Condition*. Chicago: University of Chicago Press, 1958.

Badger, Reid. *The Great American Fair: The World's Columbian Exposition and American Culture*. Chicago: Nelson Hall, 1979.

Bakhtin, Mikhail M. Ed. Michael Holquist. Trans. Caryl Emerson and Michael Holquist. *The Dialogic Imagination*. Austin, TX: University of Texas Press, 1981.

Bartlett, Richard. *Yellowstone: A Wilderness Besieged*. Tucson, AZ: University of Arizona Press, 1985.

Barrett, Cyril, and Winnifrith, Tom. Ed. *The Philosophy of Leisure*. New York: St. Martin's Press, 1989.

Baudrillard, Jean. *For a Critique of the Political Economy of the Sign*. N.p.: Telos Press, 1981.

Bennett, Mildred R. *The World of Willa Cather*. Lincoln, NE: University of Nebraska Press, 1951.

Bercovitch, Sacvan. *The American Jeremiad*. Madison, WI: University of Wisconsin Press, 1978.

Bloom, Harold. Ed. *Modern Critical Views: Willa Cather*. New York: Chelsea House Publishers, 1985.

Bromell, Nicholas K. *By the Sweat of the Brow: Literature and Labor in Antebellum America*. Chicago: University of Chicago Press, 1993.

Brown, Gillian. *Domestic Individualism: Imagining Self in Nineteenth Century America*. Berkley: University of California Press, 1990.

Chase, Alston. *Playing God in Yellowstone*. New York: Harcourt, Brace & Co., 1987.

Davidson, Cathy. *Revolution and the Word: The Rise of the Novel in America*. New York: Oxford University Press, 1986.

Dennis, Helen May. Ed. *Willa Cather and European Cultural Influences*. Lewiston: Edwin Mellen Press, 1996.

Eliade, Mircea. Trans. Willard R. Trask. *The Sacred and the Profane: The Nature of Religion*. New York: Harper & Row, 1961.

Fryer, Judith. *Felicitous Space: The Imaginative Structures of Edith Wharton and Willa Cather*. Chapel Hill, NC: University of North Carolina Press, 1986.

Gates, Paul W.. *Landlords and Tenants on the Prairie Frontier: Studies in American Land Policy*. Ithaca, NY: Cornell University Press, 1973.

Goetzmann, William H. and William N. *The West of the Imagination*. New York: W.W. Norton & Co., 1986.

Haines, Aubrey L. *The Yellowstone Story: A History of Our First National Park*. 2 vols. Yellowstone National Park: Yellowstone Library and Museum Association, 1977.

Hanley, Mark Y. *Beond a Christian Commonwealth: The Protestant Quarrel with the American Republic, 1830-1860*. Chapel Hill: Univeristy of North Carolina Press, 1994.

Harris, Barbara J.. *Beyond Her Sphere: Women and the Professions in American History*. Westport, CT: Greenwood Press, 1978.

Harvey, Sally Peltier. *Redefining the American Dream: The Novels of Willa Cather*. Rutherford, NJ: Fairleigh Dickinson Press, 1994.

Hively, Evelyn Helmick. *Sacred Fire: Willa Cather's Novel Cycle*. Lanham, MD: 1994.

Horgan, Paul. *Lamy of Santa Fé*. New York: Farrar, Strauss and Giroux, 1975.

Hyde, Anne Farrar. *An American Vision: Far Western Landscape and National Culture, 1820-1920*. New York: New York University Press, 1990.

Holt, Michael. *Political Parties and American Political Development from the Age of Jackson to the Age of Lincoln*. Baton Rouge, LA: Lousianna State University Press, 1992.

Jameson, Fredric. *The Political Unconscious: Narrative as a Socially Symbolic Act*. Ithaca, NY: Cornell University Press, 1981.

Kelley, Mary. *Private Woman, Public Stage: Literary Domesticity in Nineteenth Century America*. New York: Oxford University Press, 1984.

Kolodny, Annette. *The Lay of the Land: Metaphor as Experience and History in American Life and Letters*. Chapel Hill: University of Noth Carolina Press, 1975.

---. *The Land Before Her: Fantasy and Experience of the American Frontiers: 1630-1860*. Chapell Hill: University of North Carolina Press, 1984.

Kulikoff, Alan. *The Agrarian Origins of American Capitalism*. Charlottesville: University Press of Virginia, 1992.

Lasch, Christopher. *The True and Only Heaven: Progress and its Critics*. New York: W.W. Norton & Co., 1991.

Lee, Hermione. *Willa Cather: A Life Saved Up*. London: Virago Press, 1989.

Leverenz, David. *Manhood and the American Rennaissance*. Ithaca, NY: Cornell University Press, 1989.

Limerick, Patricia Nelson. *Legacy of Conquest: The Unbroken Past of the American* West. New York: W.W. Norton &Co., 1987.

Lukács, Georg. *The Theory of the Novel*. Cambridge, MA: MIT Press, 1971.

Marx, Karl. Trans. Martin Nicolaus. *Grundrisse: Foundations of the Critique of Political Economy*. New York: Penguin Books, 1973.

Marx, Leo. *The Machine in the Garden: Technology and the Pastoral Ideal in America*. New York: Oxford University Press, 1964.

Merchant, Carol. *The Death of Nature: Women, Ecology, and the Scientific Revolution*. New York: Harper & Row, 1980.

Miller, Perry. *Errand Into the Wilderness*. Cambridge, MA: Harvard University Press, 1956.

Myres, Sandra L.. *Westering Women and the Frontier Experience: 1800-1915*. Albuquerque, NM: University of New Mexico Press, 1982.

Nash, Roderick. *Wilderness and the American Mind*. New Haven: Yale University Press, 1967.

Novak, Barbara. *Nature and Culture: American Landscape and Painting, 1825-1875*. New York: Oxford University Press, 1980.

Osborne, William. *Caroline M. Kirkland*. New York: Twayne Publishers, Inc., 1972.

Pocock, J.G.A.. *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. Princeton, NJ: Princeton University Press, 1975.

Rosowski, Susan J. *The Voyage Perilous: Willa Cather's Romanticism*. Lincoln, NE: University of Nebraska Press, 1986.

Rothenberg, Winifred Barr. *From Market-Places to a Market Economy: The Transformation of Rural Massachussetts, 1750-1850*. Chicago: University of Chicago Press, 1992.

Schullery, Paul. *American Fly Fishing: A History*. New York: Nick Lyons Books, 1987.

Sears, John F. *Sacred Places: American Tourist Attractions in the Nineteenth Century*. New York: Oxford University Press, 1989.

Sellers, Charles. *The Market Revolution: Jacksoninan America, 1815-1846*. New York: Oxford University Press, 1991.

Silbey, Joel H. *The American Political Nation*. Stanford, CA: Stanford University Press, 1991.

Slotkin, Richard. *Regeneration Through Violence: The Mythology of the American Frontier, 1600-1860*. Middletown, CT: Wesleyan University Press, 1973.

Smith, Henry Nash. *Virgin Land: The American West as Symbol and Myth.* Cambridge, MA: Harvard University Press, 1950.

Stephanson, Anders. *Manifest Destiny: American Expansion and the Empire of Right.* New York: Hill and Wang, 1995.

Trachtenburg, Allen. *The Inorporation of America: Culture and Society in the Gilded Age.* New York: Hill and Wang, 1982.

Turner, Frederick Jackson. *The Frontier in American History.* (Huntington, NY: Robert E. Krieger Publishing Co., 1976.

Tuveson, Ernest Lee. *Redeemer Nation: The Idea of America's Millennial Role.* Chicago: University of Chicago Press, 1968.

Veblen, Thorstein. *The Theory of the Leisure Class: An Economic Study of Institutions.* New York: The Modern Library, 1934.

Volosinov, V.N. *Marxism and the Philosophy of Language.* Trans. Ladislav Matejka and I.R. Titunk. New York: Seminar Press, 1973.

White, Richard. *It's Your Misfortune and None of My Own: A New History of the American* West. Norman, OK: University of Oklahoma Press, 1991.

Whittlesey, Lee. *Death In Yellowstone: Accidents and Foolhardiness in the First National Park.* Boulder, CO: Roberts Rinehart Publishers, 1995.
---. *Yellowstone Place Names.* Helena, MT: Montana Historical Society Press, 1988.

Wilkinson, Charles F. *The Eagle Bird: Mapping a New West.* New York: Vintage Books, 1993.

Williams, Raymond. *The Country and the City.* New York: Oxford University Press, 1973.

Worster, Donald. *Under Western Skies: Nature and History in the American West.* New York: Oxford University Press: 1992.

Articles:

Bailey, Brigette. "Irving's Italian Landscapes: Skepticism and the Picturesque Aesthetic." *ESQ* 32(1) (Spring, 1986): 1-22.

Doenecke, Justus D. "Myths, Machines and Markets: The Columbian Exposition of 1893." *Journal of Popular Culture* (Winter, 1972): 535-49.

Lackey, Kris. "Eighteenth-Century Aesthetic Theory and the Nineteenth-Century Traveller in Trans-Allegheny America: F. Trollope, Dickens, Irving, and Parkman." *American Studies* 32(1) (Spring, 1991): 33-48.

McCloskey, John C.. "Jacksonian Democracy in Mrs. Kirkland's *A New Home, Who'll Follow?*" *Michigan History* XLV (December, 1961): 347-52.

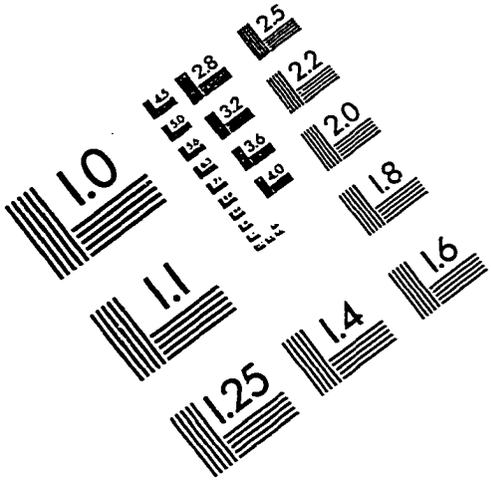
—. "Land Speculation in Michigan in 1835-36 as Described in Mrs. Kirkland's *A New Home, Who'll Follow?*" *Michigan History* XLII (March, 1958): 26-34.

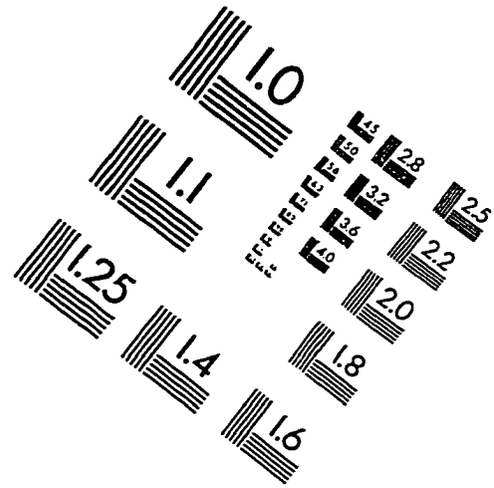
Monnett, John. Foreward. *Fly Fishing the South Platte River*. By Roger Hill. Boulder, CO: Pruett Publishing Co., 1991.

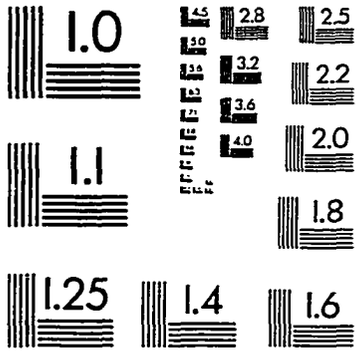
Osterud, Nancy Grey. "Gender and the Transition to Capitalism in Rural America." *Agricultural History* 67(2) (1993): 14-29.

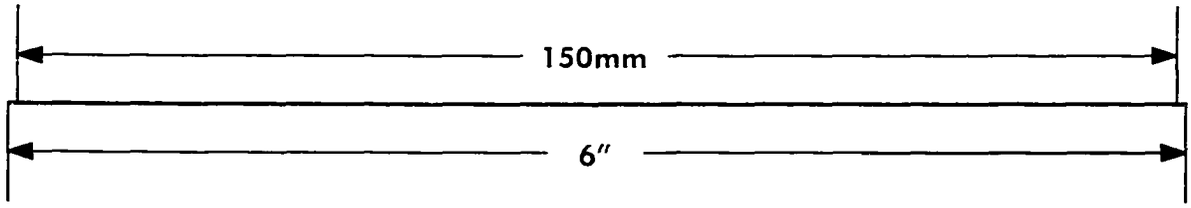
Wirth, Anne W. Fisher. "Dispossession and Redemption in the Novels of Willa Cather." *Cather Studies* vol. 1 (1990): 36-54.

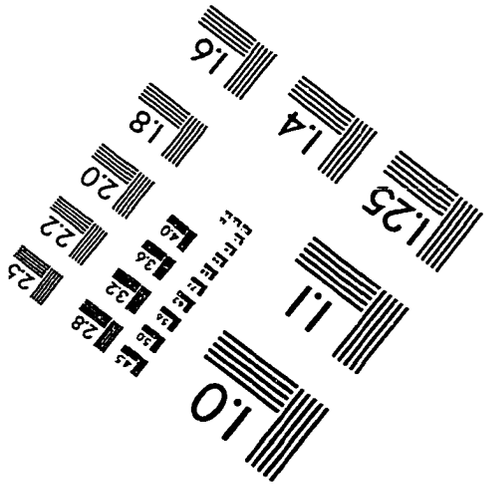
IMAGE EVALUATION TEST TARGET (QA-3)

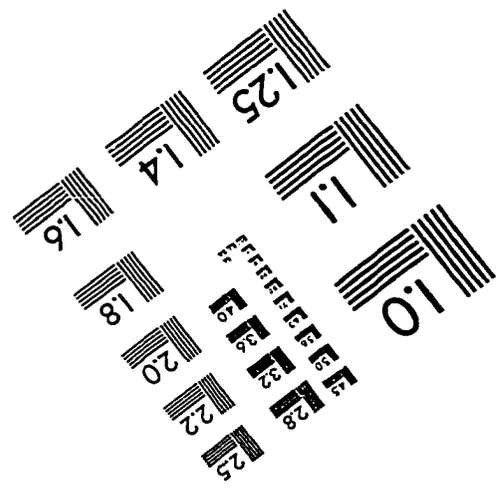












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Probing the Molecular Mechanism of Transmembrane Signaling and Kinase Regulation by the Aspartate Receptor of Bacterial Chemotaxis

by

Mark A. Danielson

B.S., University of Minnesota, 1990

A thesis submitted to the

Faculty of the Graduate School of the University of Colorado

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

1997

This thesis for the Doctor of Philosophy degree by

Mark Alan Danielson

has been approved for the

Department of Chemistry and Biochemistry

by

Joseph J. Falke

Natalie Ahn

Date 12-18-97

The final copy of this thesis has been examined by the signators, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Danielson, Mark Alan (Ph. D., Chemistry and Biochemistry)

Probing the Molecular Mechanism of Transmembrane Signaling and Kinase Regulation by the Aspartate Receptor of Bacterial Chemotaxis

Thesis directed by Professor Joseph J. Falke

The molecular mechanisms by which receptor proteins transmit information across the cellular membrane is an area of intense interest in biological science today. The aspartate receptor of bacterial chemotaxis, a useful model for the study of transmembrane signaling, is the most well characterized protein in a large family of prokaryotic proteins (with at least 68 members) displaying sequence homology (the MTP superfamily), and it shares some structural and functional similarities with eukaryotic receptors as well. The current thesis seeks to gain an understanding, on a molecular level, of the mechanisms by which the aspartate receptor transmits its signal across the membrane, and regulates the activity of the associated kinase, CheA.

In order to study the mechanism of transmembrane signaling, the water soluble ligand-binding domain of the receptor was labeled at its native Phe positions with 4-F-Phe and observed through ^{19}F NMR. The results identified the second transmembrane helix as the conduit of the aspartate-induced signal, while the first transmembrane helix was seen to be static. This aspartate-induced signal occurs rapidly, with a rate constant of ~ 10^9 $M^{-1}s^{-1}$. In addition the NMR results have shown that aspartate binding exhibits large negative cooperativity, with rapid exchange in the occupancy of the two sites. In contrast to the long-range aspartate-induced effects, formation of a non-perturbing inter-subunit disulfide bond or the binding of a non-physiological ligand produced only local effects.

To gain insight into the structure and structural mechanism of the cytoplasmic domain, the current thesis utilizes a cysteine scanning approach on a functionally important segment of the domain. This segment includes three of the receptor's four sites of regulatory methylation. A unique cysteine residue was introduced at consecutive

positions throughout this region. An assay which labeled these cysteines with a fluorescent label was able to distinguish buried from solvent-exposed residues. The solvent exposure displayed α-helical periodicity in this segment. Activity assays of the engineered receptors showed that the buried face of this helix is functionally critical. A subsequent study which assayed the activity with inter-subunit disulfides within the same set of engineered receptors showed that this helix is associated with its symmetric counterpart at the dimer interface.

Dedicated to

Barbara Fay

Danielson

and

In Loving Memory of

Gerald Duaine

Danielson

Acknowledgements

I can't possibly mention everybody who was of help to me in obtaining my degree, but a few deserve particular attention. First of all, I must thank my advisor, Joe Falke, for providing me with his guidance through all these years.

I gained a tremendous amount of aid from others in this lab. Randy Bass worked very closely with me on the methylation region study and his extensive data from *in vitro* phosphorylation assays (Figures 3.5B,C and 3.6) was a big factor in the development of the model proposed at the end of the thesis. Two undergraduates deserve special mention. Matt Coleman was incredibly helpful in carrying out membrane preps and swarm assays, and enough can't be said about the huge amount of data which Tienna Kim gathered from aspartate binding assays. Others who worked on different projects on the receptor but were nevertheless helpful to me include Scott Butler and Matt Trammell.

I must also thank Renate Griffith for all of the help she gave me on molecular modeling. The other current and former members of the lab who were helpful to me are too numerous to mention, but they all have my gratitude.

Two former members of the lab, in particular, were extremely helpful and provided me with a huge amount of guidance. Steve Chervitz was essentially my mentor on all things having to do with molecular biology. He always seemed to have an answer to my questions. Speeking of always having an answer, I found it very difficult to find a question which Olve Peersen could not answer. He was a tremendous aid on everything from computers to scientific methods to graphic design consulting.

Very special mention must be made of the departmental secretaries Maryanne Rasmussen (now retired) and Lynn Sarkis, both of whom were absolutely invaluable in making all of the hurdles I had to leap on the way to a degree surmountable.

Outside of the lab I found valuable support and intriguing conservation in the person of Sandy Parkinson, of the University of Utah.

Finally, one cannot overstate the importance of the support of a loving family. They were a tremendous comfort to me through some difficult times.

Table of Contents

Chapter 1: **The Aspartate Receptor and Bacterial Chemotaxis System of *E. coli* and *S. typhimurium*.**

Abstract 1

Introduction 2

The MTP Superfamily 7

Receptor Structure 10

Receptor Mechanism 21

Experimental Approach and Rationale 26

Chapter 2: **^{19}F NMR Maps Attractant- and Disulfide Induced Conformational Changes in the Ligand Binding Domain**

Abstract 29

Introduction 30

Materials & Methods 35

Results 39

Discussion 53

Chapter 3: **The First Methylation Segment of the Aspartate Receptor Forms a Helix-Helix Interface that Plays a Critical Role in Kinase Regulation**

Abstract 57

Introduction 58

Materials & Methods 64

Results 72

Discussion 85

Chapter 4: Modeling the Structure and the Molecular Mechanism of the Aspartate Receptor

Abstract 93

Mechanistic Model of Transmembrane Signaling 94

Structural Model of the Cytoplasmic Domain 105

Mechanistic Model for Kinase Regulation 124

Bibliography 127

List of Figures

Chapter 1

1.1 The components of histidine kinase signaling pathways 3

1.2 The chemotaxis pathway of *E. coli* and *S. typhimurium*.....6

1.3 Schematic model of the aspartate receptor.....11

1.4 Sequence alignment of the periplasmic domains of the enterobacterial chemotaxis receptors.....13

1.5 α-Helix propensity of the sequences of the periplasmic domains of the enterobacterial chemotaxis receptors 14

1.6 X-ray crystal structure of the disulfide-linked aspartate receptor....17

1.7 Coordination of aspartate by the aspartate receptor 19

1.8 Crystal structures illustrating the ligand-induced structural change in the aspartate receptor periplasmic domain.....24

Chapter 2

2.1 Structure of the apo-periplasmic ligand binding domain of the *S. typhimurium* aspartate receptor 32

2.2 Glutamate binding curves at 25°C for the oxidized N36C and N36C/F107Y ligand binding domains, obtained by monitoring intrinsic tryptophan fluorescence 42

2.3 Assignment of 4-F-Phe ^{19}F NMR resonances by site-directed mutagenesis.....43

2.4 Effect of attractant ligands on the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand-binding domain.....45

2.5 Effect of a paramagnetic probe on the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand binding domain.....45

2.6 Effect of 1,10-phenanthroline on the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand binding domain 47

2.7 Aspartate titration of the 19F NMR spectrum of the 4-F-Phe-labeled ligand binding domain.....49

2.8 Effect of the Cys36-Cys36' crosslink on the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand-binding domain 52

2.9 Schematic model of the structural and kinetic aspects of aspartate binding 56

Chapter 3

3.1 Schematic secondary structure model for the homodimeric cytoplasmic domain of the aspartate receptor.....60

3.2 Measuring solvent accessibility through thiol reactivity.....69

3.3 Cysteine scanning analysis of chemical reactivity and solvent exposure in helix α2 of the periplasmic domain.....74

3.4 Cysteine scanning analysis of chemical reactivity in the targetted region of the cytoplasmic domain 75

3.5 Effect of cysteine substitutions on receptor activity.....78

3.6 Effect of lock-on disulfides on receptor-mediated kinase regulation. 83

3.7 Model for cytoplasmic helices α5 and α6, displaying the experimentally determined solvent exposures and activity effects. 89

3.8 Model for the packing of the first methylation helices at the subunit interface.....91

Chapter 4

4.1 Ligand coordination by three different small-molecule receptors....95

4.2 Ligand coordination by two different aspartate receptors 100

4.3 Effect of aspartate on the ligand binding pocket of the aspartate receptor.....102

4.4 Long range effects of aspartate on the structure of the aspartate receptor.....103

4.5 Sequential map of the cytoplasmic domain of the bacterial chemotaxis receptors.....107

4.6 Sequence alignment of the cytoplasmic domains of the proteins of the MTP superfamily 109

4.7 α-helical propensity of the MTP superfamily as predicted by NNpredict.....117

4.8 Schematic depiction of the aspartate receptor 118

4.9 Proposed packing of the helices in the methylation region.....121

List of Tables

Chapter 1

1.1 Summary the proteins of the MTP superfamily. 9

Chapter 2

2.1 Positional parameters of phenylalanine residues in the ligand-binding domain.....34

2.2 Tyrosine substitutions at phenylalanine positions: effect on ligand binding 42

2.3 Chemical shifts of the assigned 4-F-Phe resonances of the N36C ligand binding domain.....44

Chapter One

The Aspartate Receptor and the Bacterial Chemotaxis Pathway

ABSTRACT

The aspartate receptor of *Escherichia coli* and *Salmonella typhimurium* chemotaxis is the prototype of a very large family of transmembrane receptor proteins that are ubiquitous in prokaryotes and have been discovered in eukaryotes, as well. All of these receptors control a conserved type of phosphorylation pathway initiated by a receptor-associated histidine kinase. More generally, the aspartate receptor illustrates many of the fundamental concepts underlying receptor control of kinase activity and, due to its accessibility, is a useful model for the study of transmembrane signaling. The structure of the periplasmic and transmembrane domains of the receptor is well characterized and, with the help of the ^{19}F NMR study described in this thesis, a molecular picture for how the ligand-induced signal is propogated across the membrane has been developed. A full understanding of receptor function, however, requires more knowledge than we currently possess about the structure of the cytoplasmic domain, and the mechanism by which this domain regulates the associated kinase. The present work uses a recently developed cysteine and disulfide scanning approach to analyze a functionally vital portion of the cytoplasmic domain of the intact, membrane-bound receptor: the results reveal a critical regulatory α-helix that plays a central role in the mechanism of kinase regulation.

INTRODUCTION

Bacterial cells live in precarious environments. In order to survive, they must be able to recognize changes in their environment and alter their behavior accordingly. To accomplish this, bacteria have developed elegant signaling pathways to percieve and respond adaptively to their environment. The molecular mechanisms by which bacteria accomplish this feat are the subject of intense interest in current biological science, since further knowledge will not only extend our understanding of bacterial function, but will also provide important insights into analogous eukaryotic signaling pathways. The chemotaxis system of *Escherichia coli* and *Salmonella typhimurium* is a prototypical example of signal transduction in bacteria, owing to its accessibility to genetic, biochemical, and structural approaches.

Bacterial chemotaxis is accomplished through a signaling motif termed the "two component" or "histidine/aspartate kinase" pathway (Figure 1.1). Such pathways represent the predominant mode of signaling in bacteria, and more recently, related pathways have been identified in numerous eukaryotic species (4, 23, 59, 125, 160, 191, 193, 198, 221). The "two component" term refers to the two central, highly conserved components: the histidine kinase component, or "transmitter," and the aspartate kinase component, designated the response regulator, or "reciever." In practice, however, two component pathways generally involve a third component (26, 87, 162), typically a transmembrane receptor or "sensor" that detects an environmental or physiological stimulus and regulates the histidine kinase. Specific cues recognized by these receptors may include a chemo-attractant or repellent; the presence of a host for invasion, symbiosis, or pathogenesis, changes in carbon, nitrogen, electron acceptor, or phosphate sources; changes in medium osmolarity or other physiological stress; or hormonal signals from other cells during growth or development (4, 23, 36, 57, 59, 60, 125, 135, 157, 160, 191, 193, 197, 198, 221).

Whether covalently or non-covalently, the receptor element associates with the histidine kinase to form the locus of response control. Upon detecting the initial signal, the receptor either up- or down-regulates the activity of the histidine kinase. When active, the kinase autophosporylates. Subsequently, the response regulator catalyzes the transfer of the same phosphate to a specific aspartate residue within its own active site, providing regulation that controls a cellular biochemical pathway. The response regulator also catalyzes its own dephosphorylation, although in some cases a separate phosphatase is utilized to enhance the rate of dephosphorylation or to provide an additional mode of regulation to the pathway (59, 93, 211).

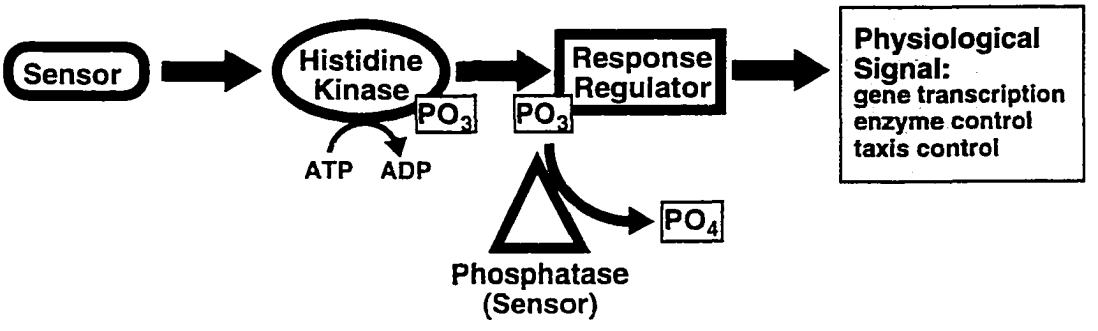


Figure 1.1. The components of histidine kinase signaling pathways. (i) The sensor protein, typically a transmembrane receptor with two membrane-spanning helices, senses environmental stimuli. (ii) The histidine kinase transmitter, regulated by the sensor, autophosphorylates. (iii) The phosphoryl group is then transferred from the histidine residue of the histidine kinase to an aspartate residue of the response regulator. (iv) Though the response regulator can catalyze its own dephosphorylation, in some pathways a separate phosphatase is used to generate more rapid dephosphorylation or to provide additional pathway regulation (59).

The chemotaxis system of *E. coli* and *S. typhimurium* (reviewed in 23, 59, 60, 135,160, 191, 193) utilizes its histidine/aspartate kinase pathway to alter the swimming behavior of the cell, enabling it to swim up a gradient of attractant or down a gradient of repellant. In the absense of a chemical gradient, the cells swim in a random three-dimensional pattern, consisting of runs of straight swimming, punctuated by tumbles during which the cell stops, reorients itself, and swims in a random new direction. When the cell is swimming up an attractant gradient or down a repellent gradient, the probability of a tumble event is decreased such that the cell exhibits long periods of smooth swimming. When the cell is swimming down an attractant gradient or up a repellent gradient, the probability of a tumble event is increased such that the cell exhibits highly tumbly behavior. (19, 23, 59, 160, 191, 193)

The molecular pathway is organized as shown in Figure 1.2. In the presence of CheA and CheW, the chemotaxis receptors form clusters, many of which are localized at one end of the cell, termed the "nose" (133, 134, 161, 184). Each transmembrane receptor exists as a ternary complex with the histidine kinase CheA and the coupling protein CheW. This complex is kinetically stable [half-time of 7 min for exchange of CheA or 17 min for exchange of CheW (71, 177, 212)] on the timescale of chemotaxis signaling [response time of 0.2 s (101, 180)]. Both the clustering and the ternary complex formation are ligand-independent (71, 133, 134, 177, 184, 212). Thus, the cell contains clusters of receptor/CheA/CheW complexes which constitute a "central command" for the chemotaxis system. Other components associate transiently with the ternary complex, including the response regulator CheY and the adaptation enzymes CheR, which binds to the C-terminal tail of the receptor (219), and CheB, which competes with CheY for binding to CheA (120).

In the absence of attractant or the presence of repellant, the chemotaxis receptors activate the histidine kinase (25, 41, 153). The kinase then transfers the phosphate to the response regulator CheY, which then diffuses to the flagellar motor (this represents the

rate-limiting step of the chemotactic response) (16). Upon reaching the motor, CheY associates with the switch protein FliM and causes the motor to switch from counter-clockwise to clockwise rotation (16, 23, 132), which results in a tumbling event. The molecular mechanism whereby phospho-CheY induces motor reversal is not known, but preliminary models exist (27, 56). Upon addition of attractant, the phosphotransfer activity of the receptor/kinase complex is downregulated over 10^2-fold (25, 41, 153). The pool of phospho-CheY is rapidly lost due to auto-dephosporylation (half-time of 6 s) (59, 86), and the rate of this degradation is further enhanced by the CheZ protein, which acts as a phosphatase (59, 86). The latter, poorly understood protein may provide as yet unknown means to regulate the pathway.

The activity of the ternary complex is also modulated by covalent methylation and demethylation by the adaptation enzymes CheR and CheB, respectively. The effect of CheR-catalyzed receptor methylation is to increase the activity of the bound histidine kinase (24, 43, 195), thus countering the effect of ligand binding. The methylesterase/ deamidase CheB contains a regulatory domain homologous to CheY (86, 130). When this domain recieves a phosphate group from CheA, the enzymatic domain is activated (130). As a result of this feedback, the steady-state level of receptor methylation is high at high concentrations of attractant and low at low concentrations of attractant. Not only does this adaptation allow the receptor to function amidst a constant background of attractant, but it also provides the receptor with a short term molecular memory (107, 148) that compares the current concentration of attractant to the concentration encountered in the immediate past. This temporal memory enables the chemotaxis system to ascertain whether the current direction of swimming is oriented up or down a spatial gradient of the receptor ligand.

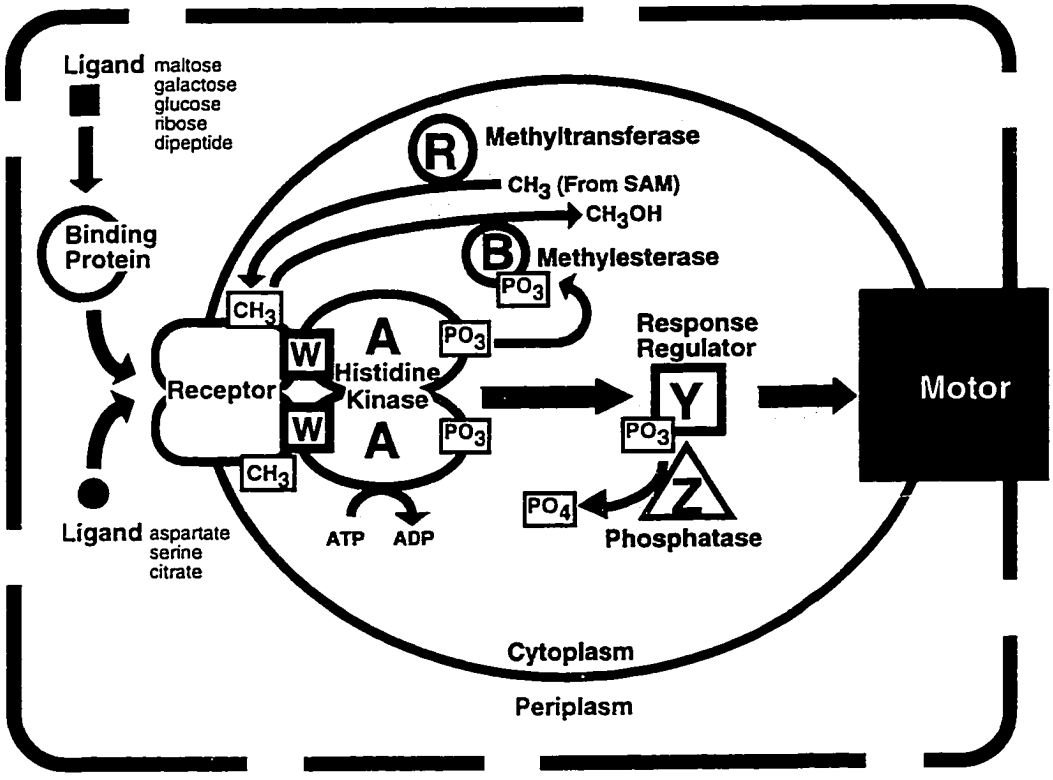


Figure 1.2. The chemotaxis pathway of *E. coli* and *S. typhimurium*. Chemosensing is initiated by the recognition of periplasmic ligands, either directly or through soluble receptor proteins. Attractant binding downregulates the activity of the associated histidine kinase CheA . When active, CheA (which forms a kinetically stable complex with the receptor and the coupling protein CheW) autophosphorylates then transfers this phosphate to the response regulator CheY, which diffuses to the motor to induce a tumbling event. The activity of the receptor/kinase complex is modulated by the methyltransferase CheR and the methylesterase CheB, which transfers a methyl group to or removes a methyl group from specific glutamate residues on the receptor, respectively. The effect of methylation is to increase the activity of the kinase. CheB is activated by phosphotransfer from CheA. Thus the steady-state level of methylation is correlated with the state of ligation. Finally, CheZ provides a specific phosphatase activity that facilitates CheY dephosphorylation. (Pathway reviewed in 23, 59, 60, 135, 160, 191, 193).

Aspartate is one of the primary chemoattractants for *E. coli* and *S. typhimurium*. Aspartate sensing is accomplished by the transmembrane aspartate receptor [sometimes called Tar, derived from the class of mutants displaying **T**axis defects to **a**spartate and **r**epellants which first identified the gene (140)], which in the case of *E. coli* also mediates the response to maltose [through the periplasmic maltose-binding protein (136)] and to certain cationic repellants (205). It is the goal of this thesis to provide insight into the mechanisms by which the receptor transmits the attractant signal across the membrane and by which this signal, together with the receptor methylation state, regulates the activity of the associated kinase.

THE MTP SUPERFAMILY

The aspartate receptor is a member of a large group of proteins displaying strong homology within the cytoplasmic domain, particularly in the region implicated in interaction with CheA and CheW. Many or all of these receptors are methylated *in vivo* as an adaptation response. We term this group the MTP (**M**ethyl-accepting **T**axis **P**roteins) superfamily, since the original nomenclature (MCP, or **M**ethyl-accepting **C**hemotaxis **P**roteins, was developed before the discovery of family members that function in phototaxis, osmotaxis, thermotaxis, etc. Table 1 lists the MTP receptors identified to date in a search of all available protein sequence databanks. [The search was performed via the program BLAST (8), using residues 290 to 470 of the *S. typhimurium* aspartate receptor as the search sequence and a cutoff value of 60. Several proteins scoring higher than 60 were discarded after a visual inspection revealed little or no conservation to the most highly conserved region (residues 357 to 419)]. The table lists 68 related proteins from twenty-five different species. In addition, eight other homologs have been identified in *Desulfovibrio vulgaris*, though not fully sequenced (54). An analysis of the transmembrane helix propensity of each sequence divides the family into four groups

based on their topology (Table 1.1). The first three groups have been identified previously from a study of phototransducers from the Archaebacterium *Halobacterium salinarium* (228). Family A, which includes the enterobacterial chemotaxis receptors, is characterized by two putative transmembrane helices separated by a periplasmic domain. This domain recognizes external ligands or other environmental conditions and transmits the signal through the transmembrane region to the highly conserved cytoplasmic domain. Family A is the largest of the four families comprising the MTP superfamily. Family B also contains two putative transmembrane helices, but instead of a significant periplasmic domain, they are connected by a short hydrophilic loop. In cases where these proteins have been functionally characterized, they were observed to associate with a different transmembrane receptor [e.g. sensory rhodopsin I or II (108, 229)]. So in this case the ligand binding domain is unnecessary and has been eliminated during evolution. Family C has considerably fewer members than the first two. This family lacks potential transmembrane helices entirely, and are thus assumed to be soluble cytoplasmic components. Though these proteins proteins have not been functionally characterized, they may associate with extrinsic transmembrane receptors or exist as independent soluble receptors for intracellular cues. Finally, several proteins, newly recognized herein, display a topology distinct from the first three families. Proteins in this family, termed Family D, contain just one segment with high transmembrane helix potential. This could mean (i) the transmembrane proteins contain just one transmembrane helix per subunit, similar to growth hormone receptors in mammalian organisms, (ii) additional, amphipathic transmembrane helices could exist that are too polar to be recognized by sequence analysis; or (iii) the lone hydrophobic region may fold to form a helical hairpin, as proposed for the aerotaxis receptor (20).

Table 1.1. Summary of the proteins of the MTP superfamily.

Name	Source Organism	Database[a]: Locus	Function[b]	Start/Stop[c] TM1	TM2	Family	Ref.
Tars	*Salmonella typhimurium*	pir:Qrebdt	Aspartate chemotaxis	7-33	191-212	A	(174)
Tare	*Escherichia coli*	swiss:Mcp2_Ecoli	Aspartate/ Maltose chemotaxis	7-33	191-212	A	(110)
Tsr	*Escherichia coli*	swiss:Mcp1_Ecoli	Serine chemotaxis	7-30	191-214	A	(29)
Tcp	*Salmonella typhimurium*	swiss:Mcpc_Salty	Citrate chemotaxis	6-32	190-213	A	(222)
Tse	*Enterobacter aerogenes*	swiss:Mcps_Entae	Serine chemotaxis	7-33	193-216	A	(48)
Tas	*Enterobacter aerogenes*	swiss:Mcpd_Entae	Aspartate chemotaxis	7-33	184-208	A	(48)
Tap	*Escherichia coli*	spswissMcp4_Ecoli	Dipeptide chemotaxis	7-33	186-210	A	(110)
Trg	*Escherichia coli*	swiss:Mcp3_Ecoli	Glucose/Galactose/ Ribose chemotaxis	17-43	197-221	A	(100)
Aer	*Escherichia coli*	swiss:Air_Ecoli	Aerotaxis	167-186	191-204	B[d]	(20)
McpArs	*Rhodobacter sphaeroides*	pir:S54262	Chemotaxis	—	64-84	D	(214)
McpArc	*Rhodobacter capsulatus*	pir:JC4735	Chemotaxis	6-26	268-288	A	(141)
McpBrc	*Rhodobacter capsulatus*	pir:JC4736	Chemotaxis	2-18	191-210	A	(141)
Orf1r	*Rhizobium meliloti*	pir:S61831	Chemotaxis	—	—	C	(77)
Orf2a	*Agrobacterium tumefaciens*	gb_ba:U60011	Unknown	—	63-80	D	(103)
Orf1c	*Clostridium thermocellum*	swiss:Cps_Clotm	Unknown	—	122-145	D	(85)
McpArl	*Rhizobium leguminosarum*	gb_U232040	Unknown	6-26	139-214	A	(30)
McpBrl	*Rhizobium leguminosarum*	gb_ba:U81828	Unknown	17-38	428-449	A	(227)
McpAc	*Caulobacter crescentus*	swiss:Mcpa_Caucr	Chemotaxis	6-29	189-214	A	(5)
McpEe	*Escherichia coli*	pir:J01475	Unknown	151-169	173-190	B	(6)
HlyB	*Vibrio cholerae*	sweiss::Hlyb_Vibch	Haemolysin secretion	10-31	199-216	A	(7)
AcfB	*Vibrio cholerae*	gb_ba:U39068	Colonization factor	7-25	279-297	A	(58)
McpAp	*Pseudomonas aeruginosa*	gb_ba:D50642	Chemotaxis	11-29	276-296	A	(113)
PilJ	*Pseudomonas aeruginosa*	swiss:Pilj_Pseae	Twitching motility	18-37	311-344	A	(53)
TcpI	*Vibrio cholerae*	swiss:Tcpi_Vibch	Toxin-coregulated-pilus gene	5-24	264-285	A	(83)
FrzCD	*Myxococcus xanthus*	swiss:Frzc_Myxxa	Aggregation	—	58-85	D	(138)
McpItm	*Thermotoga maritima*	gb_ba:E14030	Unknown	—	—	C	(104)
McpItp	*Treponema pallidum*	gb_ba:U56999	Chemotaxis	50-70	188-219	A	(80)
McpIItp	*Treponema pallidum*	gb_ba:AF016689	Chemotaxis	15-32	42-60	A	(78)
McpAtd	*Treponema denticola*	gb_ba:AF012922	Unknown	—	85-103	D	(112)
McpBtd	*Treponema denticola*	gb_ba:U84257	Unknown	—	3-23	D	(13)
McpAb	*Bacillus subtilis*	swiss:Mcpa_Bacsu	Chemotaxis	16-37	282-301	A	(81)
McpBb	*Bacillus subtilis*	swiss:Mcpb_Bacsu	Chemotaxis	17-35	284-302	A	(81)
McpCb	*Bacillus subtilis*	swiss:Mcpc_Bacsu	Chemotaxis	9-30	278-295	A	(176)
McpDb	*Bacillus subtilis*	gb_ba:D1020923	Unknown	—	—	C	(183)
McpEb	*Bacillus subtilis*	gb_ba:E325190	Unknown	—	—	C	(154)
TlpA	*Bacillus subtilis*	swiss:Tlpa_Bacsu	Chemotaxis	17-37	279-301	A	(81)
TlpB	*Bacillus subtilis*	swiss:Tlpb_Bacsu	Chemotaxis	18-35	283-301	A	(81)
TlpC	*Bacillus subtilis*	swiss:Tlpc_Bacsu	Chemotaxis	10-29	189-207	A	(82)
HtrIhh	*Halobacterium halobium*	swiss:Htri_Halha	Phototaxis	14-28	39-52	B	(224)
HtrIhs	*Halobacterium salinarum*	swiss:Htri_Halsa	Phototaxis	14-28	39-52	B	(108)
HtrIIhs	*Halobacterium salinarum*	gb_ba:U62676	Phototaxis	19-36	279-302	A	(229)
HtrIIIhs	*Halobacterium salinarum*	gb_ba:X95588	Phototaxis	—	—	C	(171)
HtrIVhs	*Halobacterium salinarum*	gb_ba:E222174	Phototaxis	35-55	325-343	A	(171)
HtrVhs	*Halobacterium salinarum*	gb_ba:E222175	Phototaxis	15-32	42-58	B	(171)

HtrVIhs	*Halobacterium salinarum*	gb_ba:E222176	Phototaxis	27-47	298-316	A	(171)
HtrAhs	*Halobacterium salinarum*	gb_ba:U785435	Phototaxis	—	—	C	(228)
HtrBhs	*Halobacterium salinarum*	gb_ba:U75436	Phototaxis	—	—	C	(228)
HtrChs	*Halobacterium salinarum*	gb_ba:U75437	Phototaxis	16-32	286-304	A	(228)
HtrDhs	*Halobacterium salinarum*	gb_ba:U75438	Phototaxis	27-47	298-316	A	(228)
HtrFhs	*Halobacterium salinarum*	gb_ba:U75439	Phototaxis	35-55	325-343	A	(228)
HtrHhs	*Halobacterium salinarum*	gb_ba:U74668	Phototaxis	—	—	C	(31)
HtrJhs	*Halobacterium salinarum*	gb_ba:U53365	Phototaxis	42-60	82-100	A	(228)
HtrIIhv	*Haloarcula vallismortis*	swiss:Htr2:Halva	Phototaxis	—	—	C	(182)
HtrIInp	*Natronobacterium pharaonis*	swiss:Htr2_Natph	Phototaxis	25-41	59-82	B	(181)
DcrA	*Desulfovibrio vulgeris*	swiss:Dcra_Desvh	Chemotaxis	11-33	188-208	A	(55)
DcrH	*Desulfovibrio vulgeris*	gb_ba:U30319	Chemotaxis	8-27	403-423	A	(54)
TlpAhp	*Helicobacter pylori*	gb_ba:AE000573	Chemotaxis	11-30	301-323	A	(204)
TlpBhp	*Helicobacter pylori*	gb_ba:AE000532	Chemotaxis	14-30	212-231	A	(204)
TlpChp	*Helicobacter pylori*	gb_ba:AE000530	Chemotaxis	14-30	212-231	A	(204)
HlyB	*Helicobacter pylori*	gb_ba:AE000573	Haemolysin secretion	—	—	C	(204)
Ant	*Helicobacter pylori*	gb_baU39068	Unknown	—	—	C	(88)
McpIs	*Synechocystis sp.*	gb_ba:D1011438	Unknown	91-109	138-159	B	(96)
McpIIs	*Synechocystis sp.*	gb_ba:D1011437	Unknown	—	—	C	(96)
McpIIIs	*Synechocystis sp.*	gb_ba:D1017932	Unknown	220-241	530-549	A	(95)
McpIVs	*Synechocystis sp.*	gb_ba:D1017761	Unknown	—	382-406	D	(95)
Y4fA	*Rhizobium sp.*	gb_ba:AE000072	Unknown	21-40	333-351	A	(66)
Y4sI	*Rhizobium sp.*	gb_ba:AE000096	Unknown	—	19-38	D	(66)
Orf2ds	*Desulfurococcus sp.*	gb_ba:U96487	Unknown	—	—	C	(186)

[a] Database abbreviations: pir=NBRF-PIR, swiss=SwissProt, gb_ba=GENBank/EMBL. [b] Physiological role or pathway in which the protein is involved as determined biochemically or genetically, or postulated based on its location in the genome. [c] First and last position in the putative transmembrane helices, predicted using the program TMpred (89). [d] The sequence of Aer yields a region of high transmembrane potential which is too long to be a single transmembrane α-helix, yet too short to give two helices. It is thus postulated to be a special case of Family B, where a hairpin turn occurs within the membrane.

RECEPTOR STRUCTURE

Figure 1.3 schematically represents the current model for the structure of the aspartate receptor. In both the presence and absence of the ligand, the receptor is a homodimer of 60 kDa subunits. The topology of each subunit consists of a periplasmic domain of approximately 150 residues separated by two transmembrane helices from a cytoplasmic domain that is approximately 300 residues long. The structure of the periplasmic domain has been shown by X-ray crystallography to consist of a homodimer of four-helix bundles, while disulfide mapping has shown that the transmembrane domain is a four-helix bundle containing two helices from each subunit. Together the periplasmic and transmembrane regions form the transmembrane signaling domain that communicates

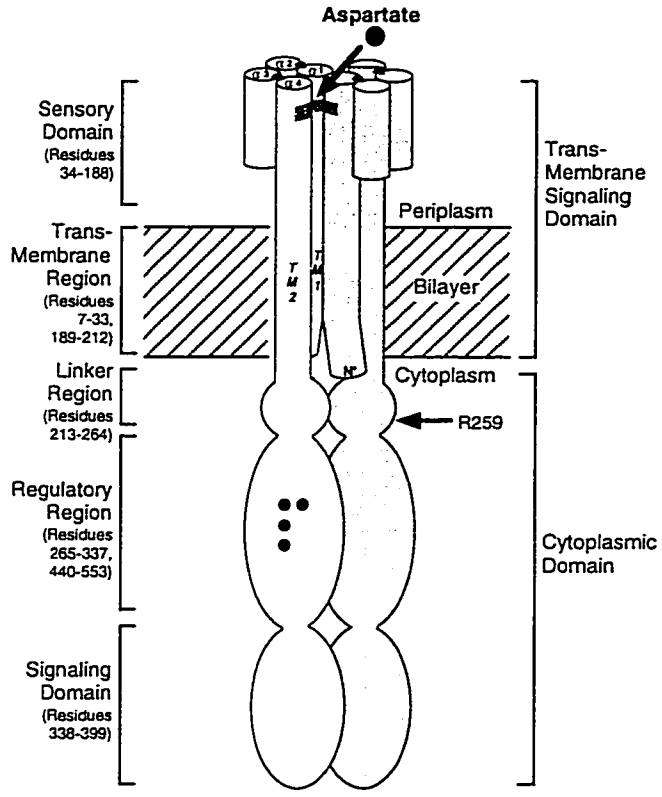


Figure 1.3. Schematic model of the aspartate receptor. The crystal structure of the periplasmic domain has revealed a dimer of four helix bundles with two symmetric aspartate binding pockets lying at the subunit interface (142). Disulfide mapping analyses indicate that the N- and C-terminal helices of the periplasmic domain continue uninterupted through the membrane, forming a four-helix bundle where the N-terminal helices associate at the dimer interface (158). Circular dichroism and hydrodynamic evidence suggests that the cytoplasmic domain packs as an elongated bundle of α-helices (124, 149, 220), but the structure of this domain has not been determined. The black circles represent the sites of regulatory methylation (E295, E302, E309, E491). Indicated on the left are the regions to which specific aspects of the receptor function have been mapped. Also indicated is the site of high proteolytic sensitivity (R259).

information from the periplasmic domain to the cytoplasmic domain. In contrast to the transmembrane signaling region, little is known about the structure of the cytoplasmic domain. As one would expect for a family of proteins designed to regulate a conserved kinase in response to a wide variety of stimuli, there is much less conservation in the periplasmic sensory domain than there is in the cytoplasmic, kinase-associated domain (compare Figures 1.4 and 4.6). In fact, a comparison of the external/periplasmic domain sequences of the 56 proteins in the MCP superfamily for which the full sequences are known suggests no general sequence conservation in this region, although the general backbone framework may be conserved.

Transmembrane Signaling Domain. Though there are no periplasmic nor transmembrane residues that are conserved throughout the superfamily, an alignment of the transmembrane signaling domains of eight enterobacterial chemoreceptors (Figure 1.4) reveals some important conserved features. First , each protein has high α-helix forming potential in the segments corresponding to the four periplasmic α-helices of the aspartate receptor, as determined by the secondary structure prediction program NNpredict (105, Figure 1.5). These putative helices are generally bordered by helix-breaking residues at locations corresponding to the inter-helical loops of the aspartate receptor. For the ribose and galactose chemoreceptor (Trg), disulfide mapping has confirmed that the periplasmic domain is a dimer of four-helix bundles, with a structure quite similar to that observed for the aspartate receptor (91, 115). Second, all of the chemoreceptors either contain a proline or a glycine at a position that aligns closely with a proline in the aspartate receptor which has been observed to generate a ligand-binding cavity in helix α4 (see below). Third, the residues found to directly coordinate the ligand in the aspartate receptor are strongly conserved among all the receptors which directly interact with small molecule attractants. The only exception is a conserved threonine that is replaced by alanine in the citrate receptor (Tcp), probably because citrate,

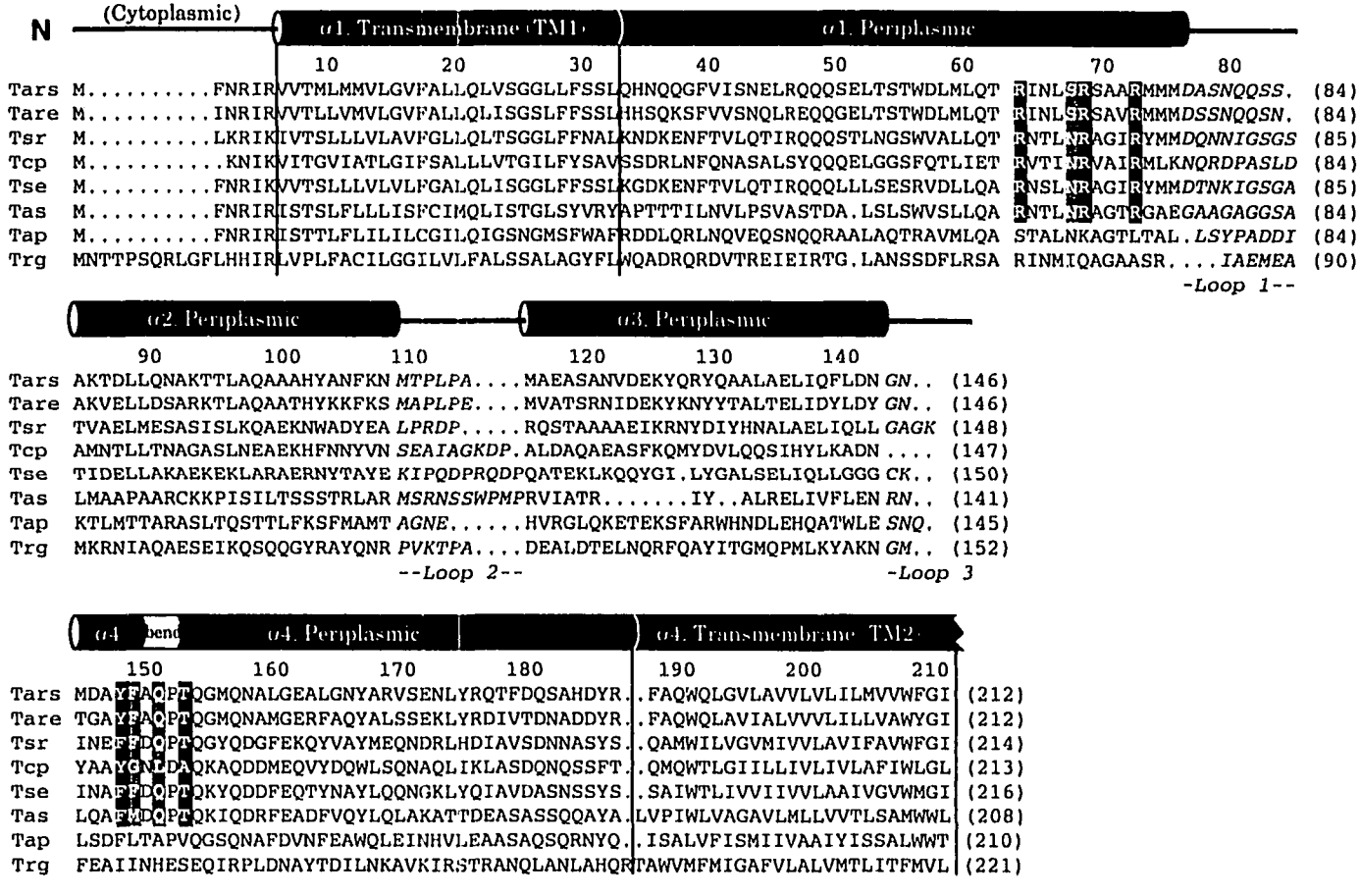
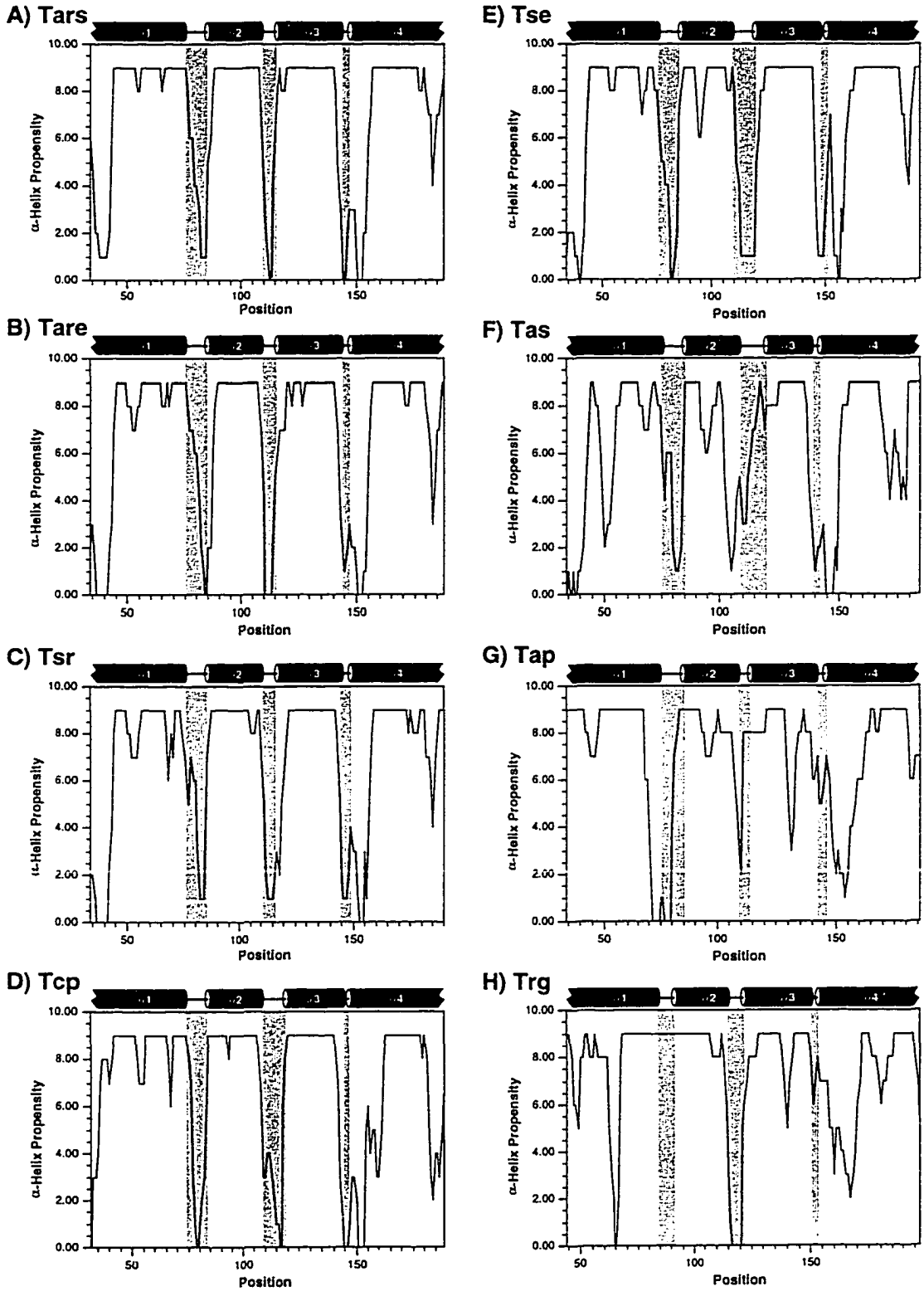


Figure 1.4. Sequence alignment of the periplasmic domains of the enterobacterial chemotaxis receptors. The cylinders above the sequences indicate the regions that are known to be α-helical in the aspartate receptor (142). The black boxes enclose the residues whose sidechains directly coordinate the ligand in the aspartate receptor (see Figure 1.6), as well as the corresponding residues in the other receptors known to directly bind a small molecule ligand. Grey boxes enclose residues which coordinate the ligand through its backbone carbonyl or a mediating water.

Figure 1.5. α-Helix propensity of the sequences of the periplasmic domains of the enterobacterial chemotaxis receptors. Propensities were determined using the program NNpredict (105). The cylinders above the graphs indicate where the α-helices of the *S. typhimurium* aspartate receptor aligns with the respective sequence (see Figure 1.4), while the shaded boxes highlight the segments where the sequence aligns with a loop.



in contrast to the amino acids, lacks the amino group which is coordinated by this sidechain.

Overall, the implications of these conserved ligand-binding site features are two-fold. (i) Each small molecule ligand interacts in a similar manner with its receptor and thus is likely to induce a similar conformational change. Consistent with this sequence-based prediction is the experimental observation of similar ligand coordination in the aspartate receptor (Tar) and the serine receptor (Tsr) as detected by X-ray crystallography and REDOR solid-state NMR, respectively (213, 226), as well as the mechanistic similarities between the aspartate receptor (Tar) and the receptor for ribose and galactose (Trg) (see *Receptor Mechanism* below). (ii) Differences in ligand specificity originate from the structural context of the binding pocket rather than from the sidechains in direct contact with the ligand. A study in which the aspartate receptor was made specific for serine by changing non-coordinating residues lends support to the importance of the surrounding context to the specificity (I Kawagishi, personal communication).

The crystal structure of the periplasmic aspartate-binding domain (residues 25 to 188) has been solved for four different states (28, 142, 225, 226): in the presence and absence of ligand, and either containing or lacking an engineered disulfide crosslink at the subunit interface. Both the disulfide-linked and the non-covalently associated dimeric fragments display an affinity for aspartate similar to that of the native receptor (145), suggesting that these isolated fragments maintain a native conformation. However, the domain lacking the engineered disulfide exhibits weaker subunit-subunit association than the detergent-solubilized intact receptor and aspartate binding of the uncrosslinked domain fails to saturate normally (145). It would thus appear that the disulfide-linked domain maintains a structure closer to native than the uncrosslinked domain; indeed, this disulfide retains normal activity in the full-length receptor (41, 62). Furthermore, Figure 1.6 shows that in the absence of the disulfide, the subunits of the periplasmic fragment spread apart as they approach what would be the transmembrane region in the intact receptor

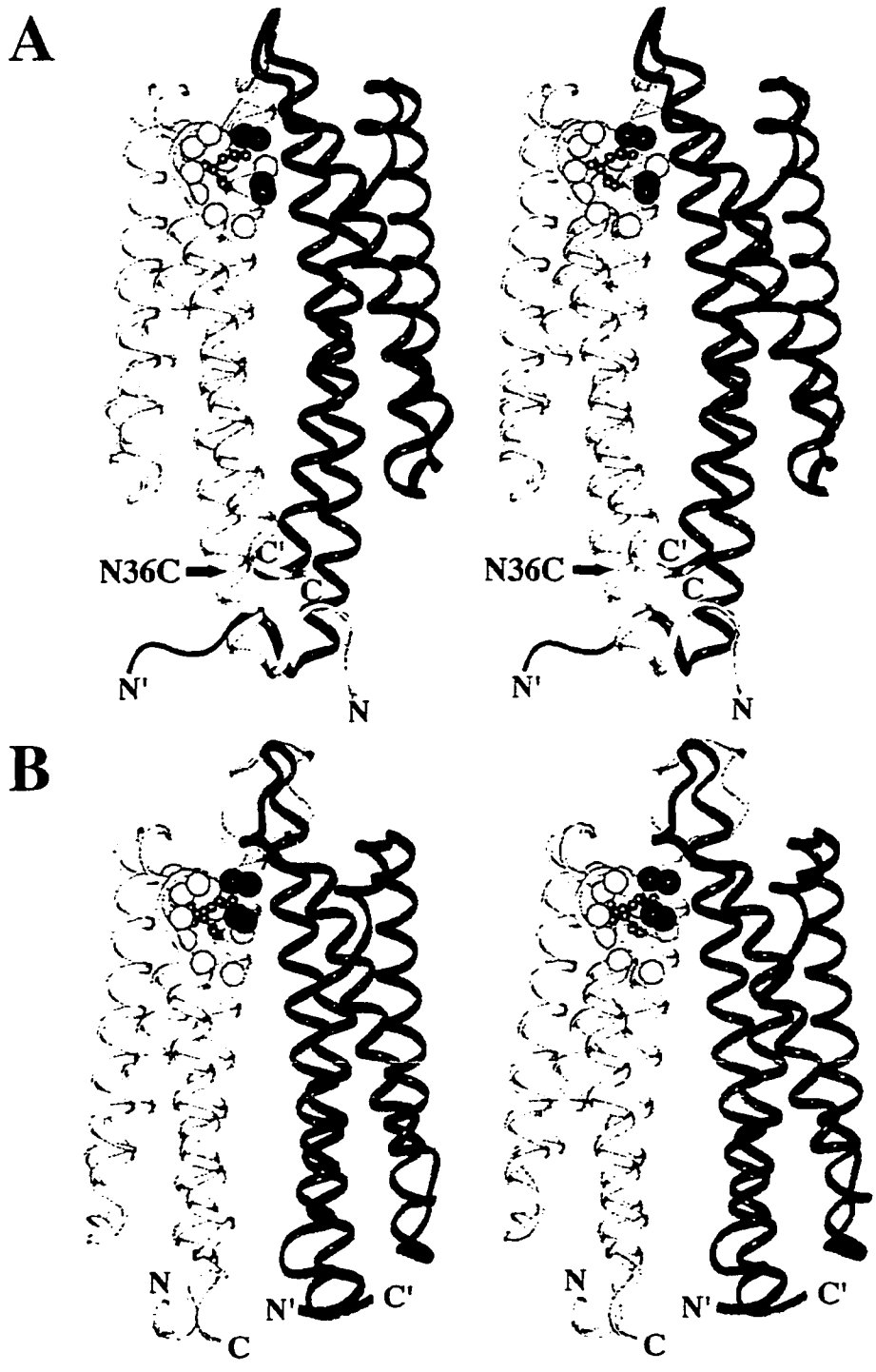


Figure 1.6. X-ray crystal structure of the disulfide-linked aspartate receptor ligand-binding domain. Shown are stereo views of the aspartate-bound receptor (**A**) with (142) and (**B**) without (225) the engineered disulfide N36C-N36'C, which lies at the dimer interface and retains normal activity in the intact receptor. Subunit A is shown in white and Subunit B is shown in black. Shown in CPK, colored according to the subunit from which they originate, are the atoms from the receptor which coordinate the bound aspartate. The bound aspartate molecule is shown in grey ball-and-stick.

(142, 225). It is therefore likely that the disulfide provides an inter-subunit stabilization that is normally maintained by contacts within the transmembrane and cytoplasmic domains.

Figure 1.6 shows the structure of the domain. Though only the aspartate-bound structures are shown in the figure, a cursory inspection of the apo structures reveal nearly identical conformations to the aspartate-bound states. A more detailed analysis reveals subtle but important differences between the structures of the two signaling states, as discussed in *Receptor Mechanism* below. The structure consists of a dimer of elongated 4-helix bundles which associate primarily through a coiled-coil interaction of the N-terminal-most α-helices. The helices are designated α1-α4 in the first subunit, and α1'-α4' in the other, proceeding from N- to C-terminal. In addition to the α1/α1' intersubunit contacts, additional contacts are observed between α1 and α4, α1 and α4', α1' and α4, as well as α1' and α4', near what would be the transmembrane region in the full-length receptor.

Aspartate binding occurs at a site distal to the membrane (Figs. 1.3 and 1.6). Association of the two subunits of the receptor creates two non-overlapping symmetrical binding sites. The interfacial nature of the binding pocket explains why aspartate binding prevents the exchange of subunits between stable dimers (143). The two binding sites exhibit negative cooperativity in direct aspartate binding measurements (22, 50), and in the crystal structure only one bound aspartate can clearly be seen (142, 225), though recent analysis raised the possibility of of partial occupation of the second site in crystals soaked with 42 mM aspartate (226). Due to its unique location both within the binding site (Figure 1.7) and at the dimer interface in the immediate vicinity of its symmetric counterpart, residue S68 was targetted in a mutagenic analysis to ascertain its role in negative cooperativity (106). A change from negative cooperativity to non-cooperative aspartate binding to positive cooperativity was achieved by mutagenically changing the identity of the residue at position 68, thus implicating this residue as an important point of

communication between the two binding sites. Since this serine residue is not conserved among the chemotaxis receptors, however (Figure 1.4), it does not appear to be the sole determinant of cooperativity. Indeed, when an asparagine is engineered into this position in the aspartate receptor, it causes positive cooperativity (106), yet the serine receptor, like most of the receptors, contains a native asparagine residue at this position, and extreme negative cooperativity is observed in serine binding to this receptor (121).

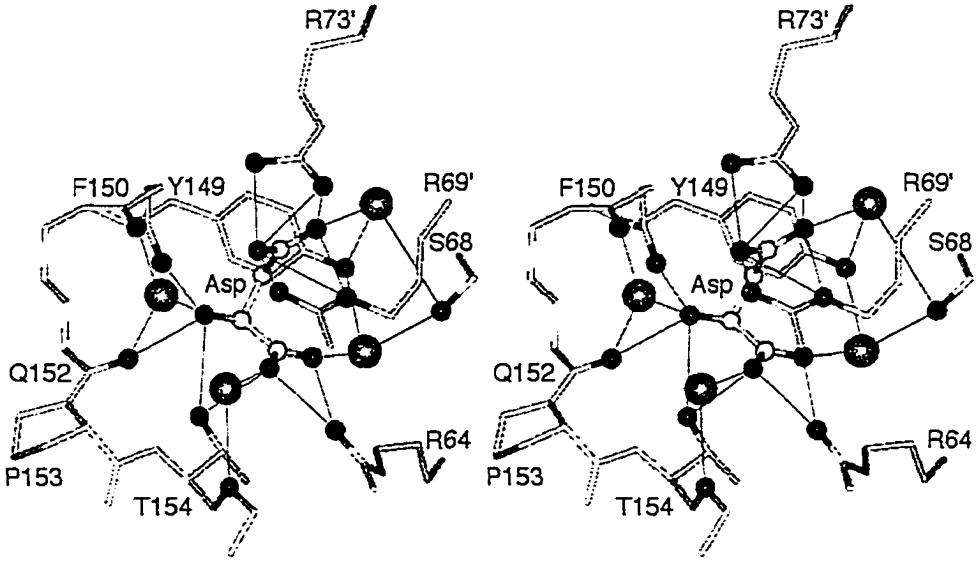


Figure 1.7. Coordination of aspartate by the aspartate receptor (226). Shown is a stereo view of the residues which directly or indirectly coordinate the ligand. Non-coordinating sidechain atoms are represented by the sticks alone, while the atoms involved in the coordination are represented by the black balls. The entire aspartate ligand is shown in ball-and-stick, with the coordinated atoms in black. The large grey balls represent bound water molecules. Hydrogen bonds or salt-bridge interactions are represented by the thin lines.

Figure 1.7 shows that the carboxyl groups of the aspartate ligand are directly or indirectly coordinated by residues from helices α1 and α1' (226), with one direct and one water-mediated contact coming from T154 of α4. The amino group is coordinated by residues within an indentation in helix α4 induced by P153. It has been demonstrated that the ligand-induced signal is transmitted through the subunit that coordinates the amino

group (70, 223), consistent with the current view of α4 as the signaling helix, as described below.

The structure of the transmembrane region has been defined by a combinatorial disulfide mapping analysis (158). The structure consists of an extension of the N- and C-terminal helices of the periplasmic domain, which continue uninterupted throughout the membrane. The subunits continue to associate primarily along the N-terminal helices (α1/TM1,α1'/TM1') while the C-terminal helices (α2/TM2,α2'/TM2') pack along the periphery to form a 4-helix bundle. These results have been verified by a random mutagenic approach (137), and studies using targetted disulfide bonds, particularly those that have systematically scanned engineered disulfides down the entire length of the major inter- and and intra-subinit helix contacts (39, 41, 61, 62, 90, 91, 115, 117). The helices α4/TM2 and α4'/TM2' ultimately reach the cytoplasmic compartment, where they are continuous with the homodimeric cytoplasmic domain (Figure 1.3).

Cytoplasmic Domain. Entering the cytoplasmic domain, the structure of the aspartate receptor (and all its relatives) becomes far more enigmatic. The water-soluble cytoplasmic domain has been isolated and has been shown to be functional (11, 124, 149). The isolated domain, however, is highly dynamic compared to typical proteins and sometimes displays heterogeneous oligomeric states (124, 179, 196). These properties have hindered attempts to develop a high-resolution structural model through NMR or crystallography. Other biophysical methods of characterization, however, have provided some insight into the structure. Circular dichroism measurements have indicated that the domain is predominantly α-helical (149, 220), while hydrodynamic studies have revealed an elongated structure (124). These results suggest a low-resolution model where the domain packs as an elongated bundle of α-helices. The domain can be isolated by proteolysis at a specific site within the linker region, lying between the second transmembrane helix and the rest of the cytoplasmic domain (149). The methylation sites

and the residues important for binding of the methyltransferase CheR display an α-helical periodicity (202, 203), suggesting that the methylation segments are α-helical, although no structural evidence is available. Sequence alignments have suggested the presence of additional specific α-helices within the domain, and little or no β-sheet (114, also Chapter 4). Finally, a region of the domain termed the signaling region has been cloned and shown to provide partial regulation of histidine kinase activity (11, 195), indicating that the signaling region is a distinct folding domain. Models for the structure of the cytoplasmic domain are discussed in Chapter 4.

RECEPTOR MECHANISM

The mechanism by which how periplasmic ligand binding and cytoplasmic methylation events modulate histidine kinase activity can be divided into two parts. (i) The transmembrane signal is the conformational change which propagates information from the binding site to the cytoplasmic domain. (ii) Kinase regulation is generated by the cytoplasmic domain, which is modulated both by the transmembrane signal and by methylation.

The Transmembrane Signal. Numerous models have been proposed for the mechanism of transmembrane signaling. Several lines of evidence have eliminated models involving a monomer-dimer equilibrium from consideration: (i) Experiments that measured the extent of disulfide formation in solubilized mixtures of receptors containing a varying ratios of engineered cysteine-containing and wildtype receptors indicated the existence of dimeric receptors both in the presence and absence of aspartate (143). (ii) The results of hydrodynamic studies of receptors in detergent solutions or in mixed-micelle systems were not consistent with a change in oligomeric state (143). (iii) Receptors covalently linked by disulfide bonds at the dimer interface were observed to transduce the normal aspartate-induced signal both through the effect of aspartate on the rate of methylation (61) and its effect on kinase activity in a reconstituted system (41).

(iv) Chemotaxis can occur *in vivo* even when the number of receptors in the cytoplasmic domain is overexpressed from high-copy plasmids (41, 62, 115, 123, 215).

Nevertheless, a model involving the aggregation of stable homodimers to form higher oligomers has been proposed. This model was based on the detection of an equilibrium between dimers and higher order oligomers of some cytoplasmic domain fragments containing mutants which, in the intact receptor, locked the receptor in an "off" state (124). Though this model has not been disproven, the failure to observe evidence supporting it, despite specific efforts by multiple laboratories to provide such evidence (12, 43, 143, 195, also Chervitz and Falke, unpublished, Bass and Falke, unpublished), suggests that the receptor must use a different mechanism of transmembrane signal transduction. The observation that chemotaxis receptors cluster to specific regions of the cellular membrane does not directly argue that dimer-dimer interactions are involved in kinase regulaton, since such interactions have recently been found to play an important role in methylation (119).

In short, transmembrane signaling is believed to involve a simple conformational change within a stable homodimer. Models for this change have been roughly divided into two groups: inter-subunit change and intra-subunit change. The former model was proposed as a result of the original analysis of the crystal structure of the ligand-binding domain in the presence and absence of aspartate. In this analysis, it appeared that the most significant difference between the structures was a slight (4°) change in orientation of the subunits (142). A subsequent analysis of the crystal structure, however, using a model-independent approach (difference-distance analysis, 40) suggested that the origional superimpositions of the apo and aspartate-occupied structures was incorrect, and that the most significant changes were localized within a single dimer.

Chapter 2 describes the first study which strongly suggested an *intra*-subunit conformational change and also provided the first evidence implicating a specific α-helix as the conduit of the transmembrane signal. In this study, a ^{19}F NMR analysis of the

isolated ligand-binding domain revealed a conformational change involving the periplasmic end of one of the two transmembrane helices (α4/TM2), while the other transmembrane helix (α4/TM1) appeared to be unperturbed. Subsequent analyses of the aspartate receptor, and the related ribose/galactose receptor supported this view and extended the conclusion to the intact, membrane-bound receptor (39, 41, 115, 117). Combining the results of the ^{19}F NMR study (Chapter 2) and disulfide-mapping results (39) with a careful reanalysis of the existing crystal structure (142), Chervitz and Falke (40) were able to develop a high-resolution picture of the molecular mechanism of the aspartate-induced transmembrane signal (Figure 1.8). The aspartate-induced changes in the crystal structure involve a 1.6 Å piston-type motion along the long axis of of one of the two α4/TM2 helices in the dimer, as well as a 5° tilting motion of the same helix. Though the significance of the tilting motion is not yet known, the significance of the piston displacement was confirmed by two independent observations: (i) Engineered disulfides that trap the extremes of the piston movement are observed to lock the receptor in either the kinase-activating (apo-like) or kinase-supressing (aspartate-bound-like) state, respectively (39). (ii) Ligand binding alters the rate of disulfide formation between engineered cysteine pairs in the ribose/galactose receptor, and the pattern of changes suggests a conserved piston displacement of α4/TM2 with this receptor as well (90).

Kinase Regulation. Though a molecular picture has now been developed for the mechanism of transmembrane signaling, the mechanism of kinase regulation is considerably less well understood, primarily because of a lack of structural information for the cytoplasmic domain. There are several aspects of kinase regulation that remain to be understood. (i) *Activation of the kinase.* Though a basal activity of the kinase in the absence of receptor can be measured in an assay with sufficient sensitivity, receptor-containing membranes stimulate this activity as much as 10^2-fold (25, 41, 153). (ii) *Down-regulation by aspartate.* Upon the addition of aspartate, the activity of the receptor-

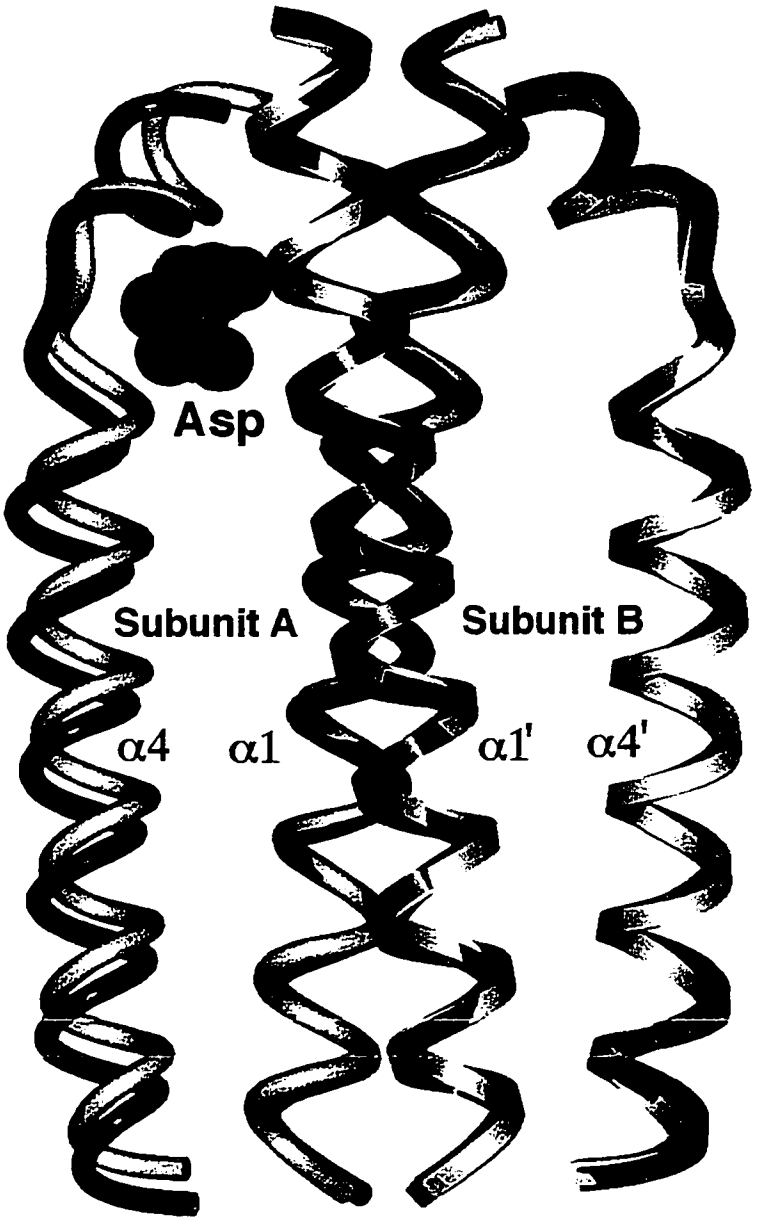


Figure 1.8. Crystal structures illustrating the ligand-induced structural change in the aspartate receptor periplasmic domain. Crystal structures were determined by Milburn, et al. (142). Shown are the two helices from each subunit which extend across the membrane. Helices α1/TM1 and α4/TM2 from subunit A are indicated by round ribbons, while the same helices from subunit B are indicated by square ribbons. The apo-receptor is shown in grey ribbon, while the aspartate-bound receptor is shown in black. The attractant aspartate molecule is shown in black CPK.

associated kinase is down-regulated at least 10^3-fold (25, 41, 153, also Bass and Falke, unpublished). (iii) *Up-regulation by methylation.* When the aspartate receptor is methylated by CheR, the activity of the associated kinase is seen to increase by as much as 10^2-fold or more (24, 43, 195). When the methylated receptor is saturated with aspartate, kinase activity returns to a level approximately the same as the unmethylated receptor without aspartate (24). Thus, the effect of methylation counteracts the effect of aspartate binding.

As was the case for transmembrane signaling, models for kinase regulation can be divided into inter- and intra-subunit mechanisms. Few specific models have been proposed. Analyses of the isolated cytoplasmic domain have suggested that dimerization of this domain is necessary for kinase activation (43, 195). Thus models have been proposed involving a switch between loosely- and tightly-associated subunits (195), or a change in the orientation of the subunits (43). It is known that CheA functions as a dimer and operates in *trans* fashion (196), thus the receptor may stabilize the proper orientation of the CheA molecules and aspartate binding could disrupt the proper orientation. The first methylation segment, and to a lesser extent the second methylation segment, is highly anionic. Thus the role of methylation in kinase activation could be to overcome steric repulsion between the subunits associating along these segments and therefore stabilize the correctly oriented dimer.

It has been suggested that modulation of kinase activity may involve a mechanism which occurs entirely within one subunit. This model has been suggested based on the observation that chemotaxis toward aspartate can occur in cells containing heterodimeric receptors which contain only one cytoplasmic domain per dimer (70, 201). Such a model is inconsistent with the *in vitro* activity assays of the cytoplasmic domain, however, demonstrating the need for dimerization for kinase activation (43, 195). Such an apparent incongruity can be rationalized by proposing that when one cytoplasmic domain of the dimer is missing, the lone cytoplasmic domains of two dimers could associate to provide

some activity. Though one would expect this activity to be lower than that with an intact dimer, one must consider that this signaling was observed *in vivo* and not *in vitro*, and in the living cell the methylation system can correct significant receptor defects via adaptation. The plausibility of inter-dimer association is supported by the observation that the chemotaxis receptors cluster within the cellular membrane (133, 134, 161, 184), and that inter-dimer interactions play an important role in receptor methylation (119).

Overall, current studies of the mechanism of the aspartate receptor generally focus on one of the two major domains. In the transmembrane signaling domain, the goal is to understand what attractant-induced conformational changes occur within the known structure to carry the signal to the cytoplasmic domain. In order to understand how the transmembrane signal or methylation is translated into kinase regulation, it will be necessary to gain a greater understanding of the structure of the cytoplasmic domain than we currently possess. These matters are addressed in Chapters 2 and 3 of the current thesis.

EXPERIMENTAL APPROACH AND RATIONALE

The present work focuses on the two major aspects of receptor function: transmembrane signaling and kinase regulation. The conformational mechanism of transmembrane signaling, which occurs in a region of well-defined structure, is probed by labeling the isolated periplasmic domain with fluorine and subsequently observing conformational changes through ^{19}F NMR spectroscopy. To study the mechanism of the modulation of kinase activity, which is regulated by a region of poorly resolved structure, the techniques of cysteine and disulfide scanning are used to probe the methylation region of the cytoplasmic domain in the full length, membrane-bound receptor.

The Transmembrane Signal: A ^{19}F NMR Study. Though crystal structures have been available for the periplasmic domain both with and without bound aspartate since 1991, the aspartate-induced structural change is small and was not

correctly characterized until 1996. Moreover, the differences between the structures are slight enough that one must be wary of the possibility that they are caused by crystal packing forces. The receptor, even the isolated periplasmic domain which is 36 kDa as a dimer, is too large for solution structure determination by current solution NMR techniques (the limit for solution structure determination by solution NMR is generally considered to be $M_r < 30$ kDa). Though the data that can be obtained through one-dimensional ^{19}F NMR is of lower resolution than that obtained through several multidimensional NMR techniques, the range of proteins accessible to the former technique is much larger ($M_r < 100$ kDa). Thus, I used ^{19}F NMR to carry out the first study of the aspartate-induced conformational change in solution using the isolated periplasmic domain fragment of the aspartate receptor.

In Chapter 2 of the current work, ^{19}F is incorporated into the *para* positions of the native phenylalanine residues of the cloned, water-soluble, ligand-binding domain fragment of the receptor, and the ^{19}F NMR spectrum of the resulting receptor is observed. After assigning resonances to specific phenylalanine positions, the effect of aspartate on the spectrum allows us to map aspartate-induced conformational changes within the three-dimensional structure of the protein. The long-range nature of attractant-induced conformational changes stands in stark contrast to the local effect of the binding of a non-physiological ligand. Additionally, the effect of forming an intersubunit disulfide bond, shown independently to leave signaling intact (41, 62), is seen to be local. A ^{19}F NMR-observed titration of the domain with aspartate reveals important information regarding the stoichiometry of aspartate binding (as discussed above). Finally, ^{19}F NMR provides important information regarding the dynamics of the periplasmic domain and its ligand-induced conformational change.

Kinase Regulation: a Cysteine and Disulfide Scanning Study. Understanding how the receptor modulates kinase activity requires a knowledge of the structure of the cytoplasmic domain of the receptor in its various signaling states.

Unfortunately, the size and dynamic nature of this domain have hindered attempts to characterize the structure through standard high-resolution techniques. Thus an alternative method of structure determination must be used The study described herein utilizes cysteine and disulfide scanning approaches to gain a greater understanding of the structure of a functionally vital segment within the cytoplasmic domain of the intact, membrane bound receptor, and its relevance in signal transduction.

Chapter 3, a collaboration study with Randal Bass, describes a study in which a unique cysteine substitution is introduced into the receptor (which contains no native cysteines) and the solvent accessibility of each substituted cysteine is measured. Application of this approach to the cytoplasmic domain reveals α-helical structural elements within the targetted region, identifies the solvent-exposed and buried faces of the helices, and pinpoints residues essential for function. Furthermore, the formation of disulfide bonds between pairs of engineered cysteines reveals a functionally critical helix-helix contact at the subunit interface of the cytoplasmic domain. Overall, the approach provides the first direct structural information for the cytoplasmic domain and also places strong constraints on the mechanism of kinase regulation.

Chapter Two

^{19}F NMR Maps Attractant- and Disulfide-Induced Conformational Changes in the Ligand Binding Domain

ABSTRACT

The isolated ligand binding domain of the chemotaxis aspartate receptor is the focus of the present study, which both a) identifies structural regions involved in the attractant-induced conformational change, and b) investigates the kinetic parameters of attractant binding. To analyze the attractant-induced conformational change within the homodimeric domain, ^{19}F NMR is used to monitor six para-fluoro-phenylalanine (4-F-Phe) positions within each identical subunit of the homodimer. The binding of one molecule of aspartate to the homodimer shifts three of the 4-F-Phe resonances significantly: 4-F-Phe150 in the attractant binding site, 4-F-Phe107 located 26 Å from the site, and 4-F-Phe180 at a distance of 40 Å from the site. Comparison of the frequency shifts triggered by aspartate and glutamate reveals that these attractants generate different conformations in the vicinity of the attractant site, but trigger indistinguishable long-range conformational effects at distant positions. This long-range conformational change is specific for attractant binding, since formation of the Cys36-Cys36' disulfide bond or the nonspecific binding of 1,10-phenanthroline to an aromatic pocket distal to the attractant site each yield a local conformational change with no detectable long-range components. The ^{19}F NMR results also reveal the association rate constant for aspartate binding ($k_{on} \sim 10^9\ M^{-1}sec^{-1}$), enabling deduction of the dissociation rate constant ($k_{off} \sim 10^3\ sec^{-1}$).

These association and dissociation rate constants indicate that the binding equilibrium and its associated conformational changes are rapid on the timescale of the chemotactic response; moreover, the observed rate constants are similar to those of the substrate sites in rapid turnover enzymes. Interestingly, only intra-subunit allosteric rearrangements are observed to be triggered by aspartate binding to the disulfide-linked homodimer, while no perturbations of the subunit interface are detected.

INTRODUCTION

The ability to alter internal functions in response to external factors such as environmental conditions and hormonal signals is essential to all cells. The primary step in a signal transduction pathway mediating the cellular response to an external cue is a transmembrane signal generated by a cell-surface receptor. The aspartate receptor of *Escherichia coli* and *Salmonella typhimurium* provides an unusually accessible model system in which to probe the nature of transmembrane signaling. This receptor belongs to a widely distributed group of prokaryotic transmembrane receptors, each of which enable a physiological response to a chemical or physical stimulus (Table 1.1; 23, 59, 60, 135, 160, 191, 193). Like the other prokaryotic receptors, the aspartate receptor is characterized by a topology possessing an external (periplasmic) ligand binding domain, a cytoplasmic signaling domain, and a pair of transmembrane α-helices connecting the two domains. Functional chimeras have been made between the aspartate receptor and the human insulin receptor (21, 147), suggesting that the aspartate receptor may share functional similarities with an even broader class of receptors.

In its physiological role, the aspartate receptor is regulated by the binding of such attractants as aspartate, aspartate analogues (42), and phenol (94) to the periplasmic ligand binding domain. The resulting transmembrane signal to the cytoplasmic domain in turn regulates a cytoplasmic phosphorylation pathway which controls the swimming behavior of the cell (for reviews see 23, 59, 60, 135, 160, 191, 193). Signaling proteins

homologous to components of the phosphorylation pathway appear to be ubiquitous in prokaryotic cells, and have recently been detected in eukaryotic cells as well (4, 36, 57, 125, 157, 197, 198, 221).

The structure of the periplasmic and transmembrane domains of the aspartate receptor has been extensively characterized. The 120 kDa receptor is a homodimer in both the presence and absence of ligand (41, 62, 143). The three-dimensional structure of the soluble periplasmic domain has been determined to 2.0 Å resolution by X-ray crystallography (28, 142, 225, 226). The architecture (Figure 2.1) consists of a dimer of four-helix bundles, with two symmetric attractant binding sites at the dimer interface. At the opposite end of the molecule, near the predicted location of the bilayer in the native receptor, an engineered Cys36-Cys36' disulfide bond covalently links the two subunits. In this same region, 3 Å from the disulfide, a 1,10-phenanthroline molecule lies nonspecifically bound in an aromatic pocket. In the intact receptor, the structure of the transmembrane domain has been characterized by targeted disulfide mapping, yielding a model for the packing of the membrane-spanning α-helices through which the transmembrane signal is communicated (61, 62, 131, 158).

The targetted disulfide approach has revealed the presence of a transmembrane conformational change triggered by attractant binding (62), but the size of the aspartate receptor and the fact that it is an integral membrane protein have hindered further structural and kinetic studies of the transmembrane signal. Even the isolated periplasmic domain, which is 36 kDa as a dimer, is too large for NMR full structure determination. One method useful in such an application is ^{19}F NMR of the protein labeled with fluorine at specific aromatic residues (reviewed in 51, 52, 72, 73, 199).

The utility of this technique stems from the inherent qualities of the ^{19}F nucleus, including (52): (i) The spin 1/2 ^{19}F nucleus occurs at 100% natural abundance and has 83% the sensitivity of ^{1}H. (ii) ^{19}F does not occur naturally in proteins; thus, there are no background signals with which to contend. (iii) Although the large anisotropy of the

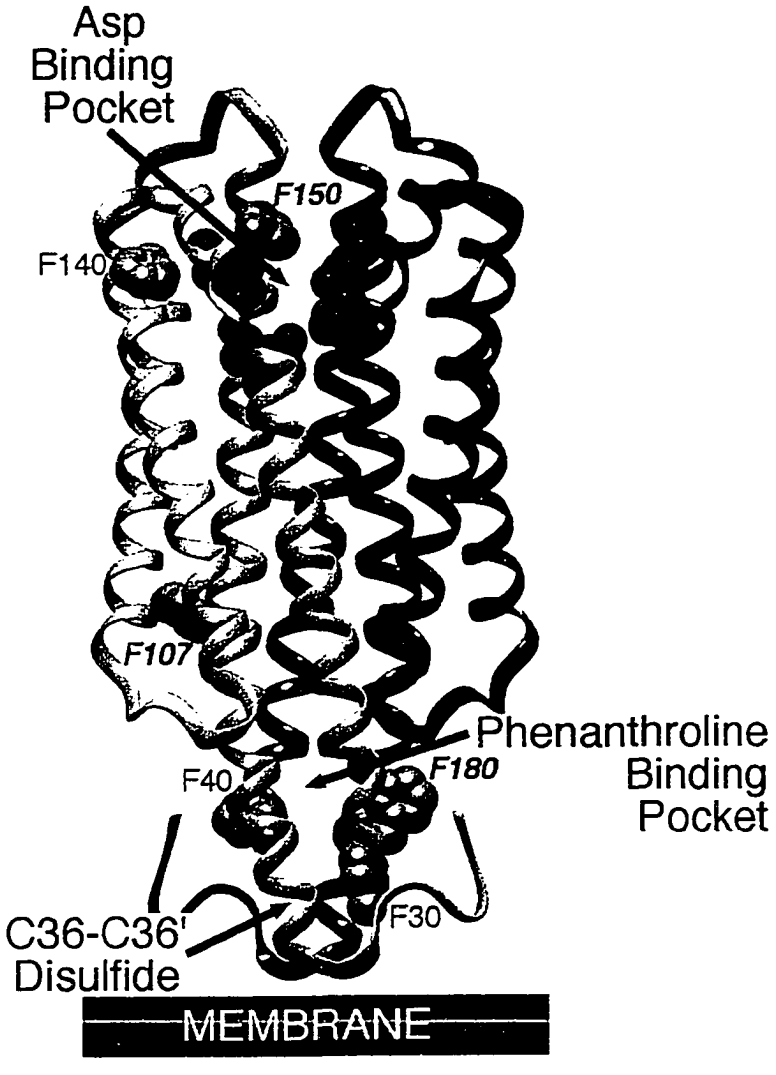


Figure 2.1. Structure of the apo-periplasmic ligand binding domain of the *S. typhimurium* aspartate receptor (43). Shown is a backbone ribbon diagram, where black and grey ribbons indicate the two different subunits. The grey CPK atoms are the Phe residues of subunit A. Shown in black are the atoms which coordinate aspartate in one of the two symmetry-related attractant binding sites. Finally, the 1,10-phenanthroline binding pocket and the Cys36-Cys36' disulfide, both located near the end of the domain distal from the attractant binding site, are indicated.

chemical shift tensor leads to broader linewidths at high field strengths, the ^{19}F chemical shift range is 100-fold larger than that of ^{1}H. This resolution, coupled with the high detection sensitivity and absence of background signals, generally yields well-resolved ^{19}F resonances in one-dimensional spectra. (iv) One-dimensional ^{19}F NMR studies generally require lower protein concentrations and shorter spectral acquisition times than do multi-dimensional NMR techniques. (v) The ^{19}F chemical shift is controlled primarily by the lone-pair electrons, which provide a large paramagnetic term in the shielding formula. The chemical shift, therefore, is exquisitely sensitive to changes in the local van der Waals environment, as well as to local electrostatic fields. (The factors affecting the chemical shift of the ^{19}F nucleus are extensively discussed in ref. 52). (vi) The exposure of specific fluorine labels to paramagnetic centers, such as a bound or aqueous metal ion, a spin-labeled analogue of a ligand, or a spin-labeled lipid probe, can be easily detected. (vii) Fluorine incorporation is generally nonperturbing, particularly when substituted for hydrogen in an amino acid sidechain. This stems from the small difference in the size of an aromatic fluorine versus an aromatic hydrogen [covalent radii of 1.35 Å vs. 1.2 Å, respectively (163)], as well as the fact that aromatic fluorine is a poor hydrogen bond acceptor and is incapable of donating a hydrogen bond (151, 163). The ability of proteins to accommodate the fluorine substitution is revealed by the fact that at least 15 proteins have been shown to be structurally or functionally unperturbed by incorporation of fluorine (14, 35, 50, 56, 68, 79, 92, 118, 126, 128, 146, 165, 166, 172, 200, 208, 217).

In addition to the ^{19}F-specific advantages, NMR methods provide unique kinetic information regarding the rate at which structural changes occur (210). The power of NMR is extended through the use of protein engineering, which allows the assignment of resonances and the resolution of ambiguities in interpretation (27, 50, 56, 164, 172).

In a previous study, 5-fluoro-tryptophan (5-F-Trp) was successfully incorporated into the intact, membrane-bound receptor and the ^{19}F NMR resonance of a lone mobile 5-

F-Trp residue was detected (63). A transmembrane conformational change was observed in that study, but resonances from multiple probe sites are needed to map out the regions of the protein participating in transmembrane signaling. In the current work, para-fluoro-phenylalanine (4-F-Phe) has been incorporated into the phenylalanine positions of the soluble ligand binding domain fragment (residues 25-188) and ^{19}F NMR spectra obtained. Each subunit possesses six phenylalanines which are fortuitously located, as illustrated in Figure 2.1 and Table 2.1, thereby providing probes in several important regions of the domain structure. Phe150 is in the immediate vicinity of the attractant binding site. Phe140 is a solvent-exposed residue near the attractant binding site. Phe107 lies in the core of a subunit 26 Å from the attractant binding site, where it provides sensitive detection of intrasubunit conformational changes. The remaining three Phe residues, Phe30, Phe40, and Phe180, are members of an aromatic cluster which surround the phenanthroline binding pocket at the dimer interface, proximal to the Cys36-Cys36' disulfide bond. Phe30 and Phe40 pack at the dimer interface and thus provide an indicator of intersubunit conformational changes. To complement the NMR data, attractant binding studies using intrinsic tryptophan fluorescence have also been carried out. Together, the results both a) place strong constraints on the cooperativity and kinetics of aspartate binding, and b) reveal regions of the ligand binding domain involved

Table 2.1. Positional Parameters of Phe Residues in the Ligand Binding Domain[a]

position	distance (Å) from bound aspartate[b]	distance (Å) from position 33[c]	distance (Å) from Cys36 disulfide[d]
30	45	9	4
40	40	13	5
107	26	29	25
140	10	55	52
150	7	57	53
180	40	16	7

[a] Calculated from the crystallographic coordinates of Milburn et al. (142). [b] Distance from the α-carbon of bound aspartate to the *para*-carbon position of the indicated phenylalanine in subunit 1. [c] Distance from the α-carbon of Leu33, near the predicted membrane interface, to the *para*-carbon position of the indicated phenylalanine in the same subunit. [d] Distance from the thiol-sulfur of Cys36 to the *para*-carbon position of the nearest indicated phenylalanine in the dimer.

in the resulting conformational change. Finally, the study addresses the effect of the intersubunit Cys36-Cys36' disulfide bond on the conformation and dynamics of the ligand binding domain.

MATERIALS AND METHODS

Materials. The expression strain used was *E. coli* RP3808 [*Δ(cheA-cheZ) DE2209 tsr-1 leuB6 his-4 eda-50 rpsL136 [thi-1 Δ(gal-attL) DE99 ara-14 lacY1 mtl-1 xyl-5 tonA31 tsx-78] /mks/*], kindly provided by Sandy Parkinson (University of Utah, 123). 4-fluoro-phenylalanine (the *para*-fluoro-isomer, 4-F-Phe) was purchased as a racemic mixture from Sigma. Glyphosate was purchased as solid tablets of Roundup Herbacide (60% glyphosate, Monsanto Laboratories). L-Aspartate and L-glutamate of >99.9% purity were purchased from Sigma.

Cloning and Mutagenesis. The gene encoding the *S. typhimurium* aspartate receptor under control of the *trc* promotor, with the first transmembrane segment replaced by the cleavable signal sequence of *E. coli* alkaline phosphatase and a stop codon inserted after residue 188, was cloned into the vector pBLUESCRIPT KS+, giving the plasmid pMK155, as previously described (145). A cysteine codon at position 36 was incorporated into the plasmid by oligonucleotide directed mutagenesis (111) to give the gene for the N36C codon in plasmid pMK155.N36C. Plasmids which express the N36C mutant bearing an additional Phe to Tyr substitution were constructed in a similar fashion.

Isolation of the Fluorine-Labeled Ligand Binding Domain. Wildtype and mutant periplasmic-domain proteins were expressed in the *E. coli* strain RP3808 bearing the appropriate version of pMK155. No induction was necessary as the *trc* promoter is constitutively expressed in this strain. The labeling medium contained 2.5 g/l bactotryptone, 1.25 g/l yeast extract, 5 g/l NaCl, and 0.75 g/l 4-fluoro-D,L-Phe. Immediately prior to innoculation, filter-sterilized stock solutions were used to yield the following final concentrations of additional components: 1 g/l glyphosate, 0.4 g/l

glucose, and 0.1 g/l ampicillin. The glyphosate served to inhibit aromatic amino acid biosynthysis (102). Cultures (500 ml per 2 l flask) were grown with vigorous aeration at 37°C for 10 hours. A 10 liter prep typically yielded 9 g of cells.

The ligand-binding domain was purified using a variation of a gentle osmotic shock procedure previously described (145). Cells were harvested from media by centrifugation [Beckman JA-10 rotor, 6,000 rpm (6,400 x g), 5 min] and washed twice in a buffer containing 10 mM Tris, pH 7.4 with HCl, 30 mM NaCl, and 0.5 mM EDTA (harvested by centrifugation after each wash). The cells were then resuspended in ice-cold spheroplast buffer containing 100 mM Tris, pH 8.0 with HCl, 500 mM sucrose, 0.5 mM EDTA, and 0.2 mM PMSF incubated at room temperature for 10 minutes. Finally, the cells were harvested by centrifugation [Beckman JA-14 rotor, 10,000 rpm (15,000 x g), 10 min] and gently resuspended in ice-cold distilled water to lyse the outer membrane. The water contained 1 mM $MgCl_2$ to maintain integrity of of the cytoplasmic membrane. The suspension was incubated on ice for 15 min. The spheroplasted cells were removed by centrifugation [Beckman JA-14 rotor, 12,000 rpm (22,000 x g), 10 min]. To the supernatant were added the following final concentrations of constituents: 50 mM Tris, pH 7.4 with HCl, 2 mM EDTA, 2 mM 1,10-phenanthroline, and 0.2 mM PMSF. The protein was recovered by ammonium sulfate precipitation. In this step, 0.26 g of solid ammonium sulfate was added per ml of starting solution. After stirring at 4°C for 1 hour, the precipitated protein was collected by centrifugation [Beckman JA-14 rotor, 12,000 rpm (22,000 x g), 30 min] and resuspended in 10 ml of 10 mM Tris, pH 8.0 with HCl. The protein was dialyzed overnight against 1 liter of 10 mM Tris, pH 8.0 with HCl, including at least two exchanges with fresh buffer.

The dialyzed sample was loaded onto a 75 ml Q-Sepharose column (2.5 cm diameter) equilibrated with 10 mM Tris, pH 8.0 with HCl. A gradient was applied of 0 to 300 mM NaCl across a total of 600 ml of 10 mM Tris, pH 8.0 with HCl: the ligand binding domain eluted between 100-150 mM NaCl. Fractions collected during the

gradient were analyzed by monitoring the absorbance at 280 nm. The peak fractions were pooled and concentrated by ultrafiltration (Amicon, YM10 membrane) to a volume of ~10 ml. The pool was then dialyzed for 2-3 days against 1 liter of 10 mM Tris, pH 8.0 with HCl, 50 mM NaCl, 50 mM KCl, and 1 mM $MgCl_2$, including at least 3 changes of buffer. A typical 10 liter prep yielded between 10 and 30 mg of protein of 85% purity. This procedure yielded the oxidized homodimer crosslinked by the Cys36-Cys36' disulfide bond; for reduced N36C domain the sample was incubated overnight under a nitrogen atmosphere, with 50 mM DTT and 0.5 mM EDTA at 4°C.

Measurement of the Concentration of the Purified Ligand-Binding Domain. The total protein concentration was determined using a bicinchoninic acid (BCA) colorimetric assay (194). To improve accuracy and precision, sodium dodecyl sulfate (SDS) was added to the reaction mixture to a final concentration of 0.1% (w/v). The resulting concentration was corrected for the purity of the ligand binding domain, as determined by SDS-PAGE (15% acrylamide gel).

In some cases the concentration of the ligand binding domain was determined by measuring the absorbance of the sample at 280 nm. An extinction coefficient for the ligand binding domain was determined by the method of Gill and von Hippel (74), yielding ε_{280} =29,200 $M^{-1}cm^{-1}$ for the wildtype dimer, ε_{280} = 29,400 $M^{-1}cm^{-1}$ for the disulfide-linked dimer, and ε_{280} = 32,000 $M^{-1}cm^{-1}$ for the tyrosine replacement mutants of the disulfide-linked dimer. This method gave a calculated ligand binding domain concentration which agreed, to within an error of 10%, with the value determined using the BCA assay.

Measurement of Ligand Binding. The binding of aspartate or glutamate to the ligand binding domain was monitored at 25°C by the resulting increase in intrinsic tryptophan steady-state fluorescence using an SLM 48000S spectrofluorimeter (λ_{ex}=283 nm, 4 nm bandwidth; λ_{em}=350 nm, 8 nm bandwidth). Samples contained 2.5 mM dimeric ligand binding domain for determination of the glutamate dissociation constant

(K_D) and 0.1 mM dimeric ligand binding domain for determination of the aspartate K_D. The low-affinity ligand glutamate was routinely used to measure the binding constants of the replacement mutants because its larger K_D eliminated the need to correct the total ligand concentration for bound ligand, as was required for aspartate binding. To eliminate the possibility of effects due to a shifting monomer-dimer equilibrium, the disulfide-crosslinked N36C domain was used in all cases. Though the emission maximum is at 336 nm, the greatest change upon the binding of aspartate or glutamate was observed at an emission wavelength of 350 nm. Non-linear least squares analysis was used to determine K_D from the untransformed binding titration:

$$F = \frac{F_0 + F_{max}[L]}{([L]+K_D)} \quad (1)$$

where F is the observed fluorescence intensity, F_0 is the fluorescence intensity before addition of ligand, F_{max} is the fluorescence change induced by saturating ligand, and [L] is the free ligand concentration.

^{19}F NMR Measurements. ^{19}F NMR spectra were obtained at 470 MHz on a Varian VXR 500 spectrometer equipped with a 5-mm ^{1}H/^{19}F probe. Samples contained 0.3-1.5 mM dimeric ligand binding domain in 10 mM Tris, pH 8.0 with HCl, 50 mM NaCl, 50 mM KCl, and 1 mM $MgCl_2$. Samples were prepared by concentrating, using ultrafiltration (Amicon, YM10 membrane), the final dialysate from the purification to a volume of 600 ml, then adding D_2O and 5-fluoro-tryptophan, the latter as an internal frequency standard (referenced to -49.5 ppm, its known chemical shift relative to TFA at 0 ppm), thereby enabling direct comparison of chemical shifts in different spectra. Standard undecoupled spectral parameters were as follows: 12,000-Hz spectral width, 16K data points, 80° pulse width, 0.68 sec aquisition time, 1.2 sec relaxation delay, 20 Hz line broadening, and temperature control at 25°C. For quantitative integrations, spectra were obtained with the pulse width decreased to 60° and the relaxation delay increased to 5.0 sec. T_1 relaxation measurements, utilizing a 180°-t-90° pulse sequence

and a 4.0 sec relaxation delay, yielded T_1 values ranging from 0.7 to 1.4 sec, within the typical range observed for 4-F-Phe resonances (56)

Determination of the Extent of Fluorine Incorporation. The ^{19}F NMR spectrum of the apo-domain was integrated, yielding four peaks of equal intensity and one of double intensity (the result of two overlapping resonances). The average intensity of the individual 4-F-Phe resonances was then compared to that of the internal standard of known concentration and this information, together with the known protein concentration, provided the extent of fluorine incorporation.

Molecular Graphics. Crystallographic coordinates of the apo- and aspartate-occupied conformations of the ligand binding domain were graciously supplied by Kim and co-workers (142), and were visualized using Biosym Technologies Insight II graphics software running on a Silicon Graphics Personal Iris workstation. Surface accessibility calculations were carried out using the program "Accessibility" developed by Handschumacher and Richards (170).

Error Estimates. Integrations of individual ^{19}F NMR resonances exhibited a relative error of ±10%, determined by comparing the relative integrals from four independent spectra. The standard deviation inherent in measurements of ^{19}F frequency shifts is ±0.1 ppm, ascertained by comparing the chemical shifts of identical resonances in four independent spectra. Errors for parameters obtained in binding curve fits were determined by nonlinear least-squares error analysis.

RESULTS

Affinity of the Ligand Binding Domain for Aspartate. In order to investigate the structural integrity of the ligand binding domain, the affinity of the oxidized N36C domain for the attractant aspartate was compared to that of the intact receptor. The affinity of the soluble domain for aspartate was measured via titration of intrinsic tryptophan fluorescence, yielding K_D=1.4±0.5 μM, a value within the range determined

for the intact wildtype receptor in the membrane (K_D=0.1 to 6 μM; ref. 22, 42), as well as for the isolated domain (K_D= 1 to 2 μM; ref. 145). This result indicates that aspartate binds to the isolated domain and triggers an essentially native conformational change, and that the engineered Cys36-Cys36' inter-subunit disulfide bond is relatively non-perturbing. Additional evidence that the engineered disulfide is non-perturbing is provided by its minimal effect on transmembrane signaling in the intact receptor (41, 61, 62, 143).

Incorporation of 4-F-Phe. The ligand binding domain was fluorine-labeled for ^{19}F NMR studies by overexpression of the cloned gene in the presence of glyphosate, which supresses the cell's ability to synthesize aromatic amino acids (102). A 4:1 ratio of *para*-fluoro-L-phenylalanine (4-F-Phe) to unlabeled L-phenylalanine was added to the growth medium to optimize production of fluorine-labeled protein. Since *E. coli* possesses the ability to discriminate against substitution at the para position of the Phe ring, the resulting efficiency of 4-F-Phe labeling was 7%. Thus, at any given Phe position within the fluorine-labeled ligand binding domain, there was a 7% probability of finding a 4-F-Phe residue.

Due to the low efficiency of fluorine labeling it was not possible to test directly the effect of the fluorine incorporation on receptor activity; however the effect of a more extreme perturbation was accessible to quantitation. In protein structures, hydroxyl substitution for aromatic hydrogen is expected to be more perturbing than fluorine substitution: this stems from the ~50% greater volume of the hydroxyl group relative to fluorine and the propensity of hydroxyl to form hydrogen bonds, while aromatic fluorine is a poor hydrogen bond acceptor and is incapable of donating a hydrogen bond (151, 163). To implement the test a set of six mutant proteins was generated, each possessing a single tyrosine substitution at a different phenylalanine position. These mutants enabled estimation of the maximum perturbation due to fluorine incorporation at the *para* position

of each phenylalanine residue. The attractant affinities of the resulting mutants were compared using a fluorescence assay to monitor glutamate binding (see Methods).

Figure 2.2 compares the glutamate binding curves, monitored by intrinsic tryptophan fluorescence, of the disulfide-linked ligand binding domain (N36C) and a representative Phe to Tyr replacement mutant (N36C/F107Y). The curves are indistinguishable. Table 2.2 shows the calculated K_D's of each replacement mutant for glutamate. The affinities of the mutant proteins for glutamate range from 4-fold tighter to 1.6-fold weaker binding. These minor effects suggest that hydroxyl substitutions at the para-position of each phenylalanine residue cause little or no perturbation of the domain. Therefore the even smaller perturbations due to 4-F-Phe incorporation are likely to be negligible.

Assignment of the ^{19}F NMR Resonances to Specific 4-F-Phe Residues. Figure 2.3A presents the ^{19}F NMR spectrum of 4-F-Phe-labeled N36C ligand binding domain in the apo-conformation. Six resonances are discernible, two of which overlap to a large degree. The observation of only six resonances from the homodimer indicates that the structures of the two subunits are the same *on average*, such that the six 4-F-Phe residues in subunit A are equivalent to the corresponding residues in subunit B. Thus the two subunits rapidly sample the same range of conformations, yielding fully overlapping spectra (see below). Integration reveals that the 4-F-Phe resonances each possess the same intensity, verifying that the fluorine labeling efficiency is the same at all six phenylalanine positions of each subunit.

The six resonances were assigned to specific 4-F-Phe residues by two methods. Where possible, direct replacement of individual Phe residues was used. In this method, a mutant bearing a Tyr substitution at a single Phe position is labeled with 4-F-Phe at the remaining Phe positions and used to generate a ^{19}F NMR spectrum, in which the ^{19}F NMR resonance from the targetted position disappears. As an example, Figure 3B illustrates the ^{19}F NMR spectrum of a mutant bearing a Tyr substitution at the Phe180

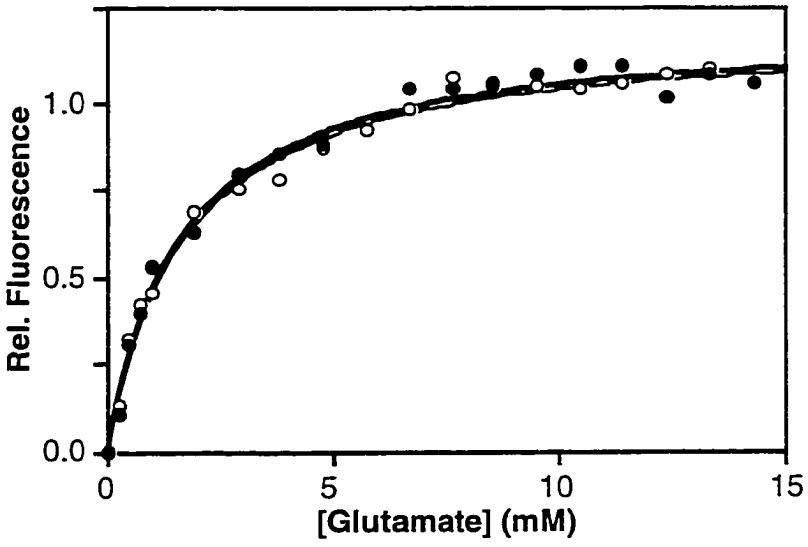


Figure 2.2. Glutamate binding curves at 25°C for the oxidized N36C (closed circles) and N36C/F107Y (open circles) ligand binding domains, obtained by monitoring intrinsic tryptophan fluorescence. A nonlinear least square best-fit curve generated for a homogenous population of sites is shown for each domain (N36C=bold line, N36C/F107Y=narrow line, best fit K_D values summarized in Table 2). The buffer was 10 mM Tris, pH 8.0 with HCl, 50 mM NaCl, 50 mM KCl, and 1 mM $MgCl_2$. The concentration of the dimeric domain was 3 μM (or 6 μM monomer).

Table 2.2. Tyrosine Substitutions at Phenylalanine Positions: Effect on Ligand binding

mutant	glutamate Kd (mM)[a]
N36C	2.0 ± 0.2
N36C/F30Y	3.2 ± 0.3
N36C/F40Y	2.2 ± 0.4
N36C/F107Y	1.4 ± 0.1
N36C/F140Y	2.2 ± 0.6
N36C/F150Y	0.5 ± 0.3
N36C/F180Y	1.3 ± 0.8

[a] Measured by monitoring intrinsic tryptophan fluorescence during glutamate titration at 25°C in 10 mM Tris, pH 8.0 with HCl, 50 mM NaCl, 50 mM KCl, and 1 mM $MgCl_2$. The concentration of the dimeric domain was 2.5 μM.

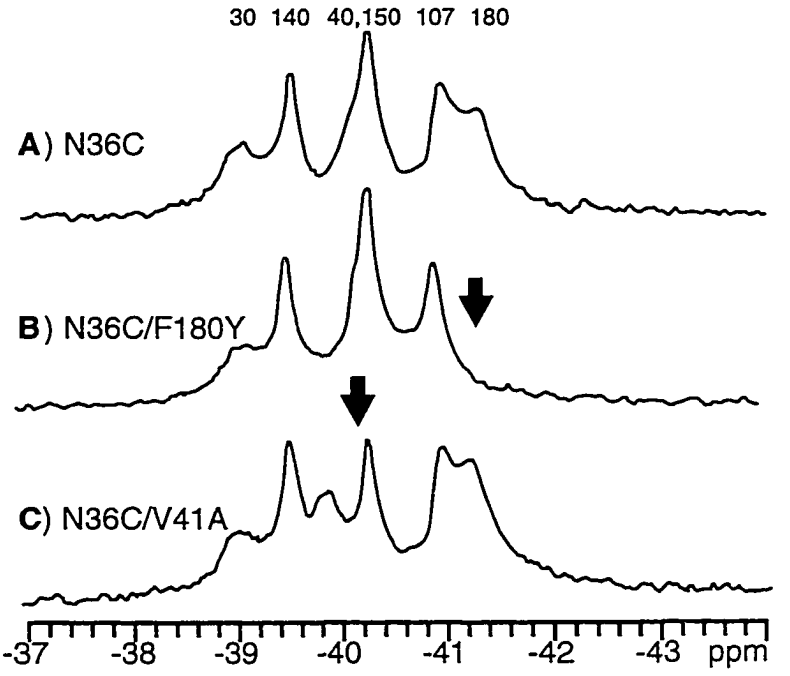


Figure 2.3. Assignment of 4-F-Phe ^{19}F NMR resonances by site directed mutagenesis. Shown are the ^{19}F NMR spectra of two 4-F-Phe-labeled domains: (**A**) N36C (**B**) N36C/F180Y, and (**C**) N36C/V41A. The arrows indicate the resonance deleted (F180) or nudged (F40) by the mutation. The final assignments provided by replacement and nudge mutational analysis are indicated (see text). Spectra were obtained at 470 MHz and 25°C in the same buffer as in Figure 2, with the addition of 10% D_2O and 50 mM 5-F-Trp. The concentration of dimeric domain was 0.6 mM.

position (N36C/F180Y); here five resonances are unchanged while the sixth is obliterated. The missing resonance is directly assigned to the 4-F-Phe180 residue. This procedure was systematically repeated for the other phenylalanine positions to complete the majority of the assignments in the spectra of different ligation and oxidation states.

In several cases, an additional method was needed to achieve unambiguous assignment. This method utilized a "nudge" mutation, wherein a residue in van der Waals contact with a single Phe residue is identified in the crystal structure and then substituted with a residue of different size (56). When the chemical shift of a specific resonance was

altered, this resonance was directly assigned to the nudged 4-F-Phe residue, as illustrated in Figure 2.3C

Together these two assignment methods allowed the unambiguous assignment of 36 ^{19}F NMR resonances associated with different ligation and oxidation states of the 4-F-Phe labeled N36C domains (Table 2.3). Of these resonance assignments, 30 were obtained by direct replacement, while 6 required nudge mutational analysis. The assignment of one resonance, 4-F-Phe140, was independently confirmed by paramagnetic broadening studies. The 4-F-Phe140 residue is the only one of the six phenylalanine positions in the crystal structure which is accessible to a solvent probe the size of Gd(III)-EDTA [4.8 Å radius (127)]. As expected, the resonance assigned to 4-F-Phe140 yielded the largest resonance broadening upon addition of the aqueous probe Gd(III)-EDTA to either the apo- or aspartate-occupied conformations (Figure 2.5).

Table 2.3. Chemical shifts of the assigned 4-F-Phe resonances of the N36C ligand-binding domain

	apo oxidized[a]	(+)Asp oxidized	(+)Glu oxidized	(+)Asp(+)Phe oxidized	(+)Asp reduced[b]	(+) Glu reduced
F30Y	-39.0	-39.0	-39.0	-38.8	-40.4	-40.4
F40Y	-40.2	-40.2	-40.2	-40.0	-40.6	-40.6
F107Y	-40.9	-41.2	-41.2	-41.2	-41.2	-41.2
F140Y	-39.5	-39.5	-39.5	-39.5	-39.5	-39.5
F150Y	-40.2	-38.9	-39.5	-38.9	-38.9	-38.9
F180Y	-41.3	-41.2	-41.2	-41.2	-40.9	-40.9

[a] Fully oxidized to form an intersubunit N36C-N36'C disulfide bond. [b]N36C and N36'C are completely reduced to free thiols.

Conformational Effects of Attractant Binding. The addition of aspartate or glutamate, both of which act as attractants through the aspartate transducer (42), induces a structural change in the oxidized N36C domain which was detected by ^{19}F NMR, as shown in Figure 2.4A-C and Table 2.3. Significant attractant-induced chemical shift changes are observed in these spectra for the 4-F-Phe150, 4-F-Phe107, and 4-F-Phe180 resonances (see Figure 1). The 4-F-Phe150 resonance exhibits different chemical

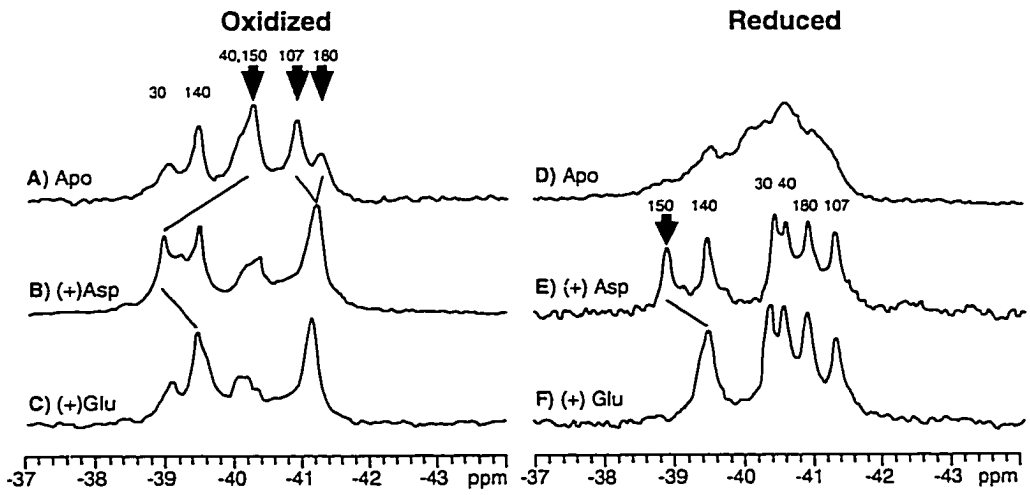


Figure 2.4. Effect of attractant ligands on the ^{19}F NMR spectrum of the 4-F-Phe labeled ligand binding domain. Shown are the spectra of the (**A**, **D**) apo, aspartate (**B**, **E**) bound, and (**C**, **F**) glutamate-bound states of the oxidized and reduced N36C ligand binding domain. The bold arrows indicate which of the assigned resonances are shifted by attractant binding; the new positions of these resonances are indicated by the light diagonal lines. Sample conditions were as in Figure 2.3; where indicated, 5 mM aspartate or 25 mM glutamate were also present. The concentration of dimeric domain was 0.7-2 mM.

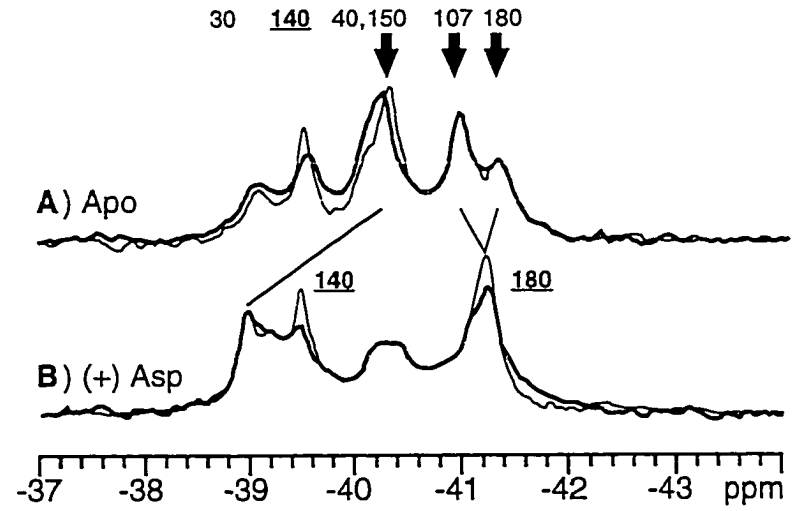


Figure 2.5. Effect of a paramagnetic probe on the ^{19}F NMR spectrum of the 4-F-Phe labeled ligand binding domain. Shown are the spectra of the the (**A**) apo- and (**B**) aspartate-bound receptor in the absence (thin line) and presence (thick line) of the paramagnetic probe $Gd^{3+}EDTA^{4-}$. Arrows and lines indicate the resonances shifted by aspartate. The resonances broadened by the probe are indicated in bold and underlined. Sample conditions were as in Figure 2.3.

shifts in the apo-, aspartate-bound, and glutamate-bound states, consistent with this residue being adjacent to the attractant binding site. Most importantly, aspartate and glutamate each cause the same chemical shift change at the 4-F-Phe107 position (-0.3 ppm), as well as at the 4-F-Phe180 position (+0.1 ppm), suggesting that these attractants generate a similar long-range conformational change. Similarly, when attractant is added to the reduced N36C domain, the 4-F-Phe150 resonances of the aspartate- and glutamate-bound states differ by 0.5 ppm, but the 4-F-Phe107 and 4-F-Phe180 resonances are shifted by either attractant the same degree from the apo-state (Figure 4D-F and Table 2.3). Together, these results suggest that aspartate and glutamate generate different local conformations at the attractant binding site, but yield the same long-range conformational change detected at the 4-F-Phe107 and 4-F-Phe180 positions. Despite its long-range nature, this attractant-induced conformational change remains localized within specific regions of the molecule, since conformational changes are not detected at the 4-F-Phe30, 4-F-Phe40, and 4-F-Phe140 positions (Figure 2.4A-C and Table 2.3).

Additional evidence for a long-range attractant-induced conformational change was provided by paramagnetic broadening studies. The 4-F-Phe180 residue, which lies 45 Å away from the attractant binding site, yielded significant resonance broadening by aqueous $Gd^{3+}EDTA^{4-}$ only in the attractant-occupied conformation of the oxidized N36C domain (Figure 2.5). Thus attractant binding significantly increases the solvent exposure of this residue.

The binding of 1,10-phenanthroline to the oxidized N36C domain was also studied as a control, since this ligand is neither an attractant nor repellent in chemotaxis (Chervitz and Falke, unpublished). In contrast to the long-range effects of aspartate or glutamate binding, 1,10-phenanthroline binding was observed to cause a more localized structural change. The effect of 1,10-phenanthroline on the ^{19}F NMR spectrum of the apo-domain is summarized in Figure 2.6A,B. As anticipated, the 4-F-Phe30, 4-F-Phe40, and 4-F-Phe180 resonances, which originate from the 1,10-phenanthroline binding site

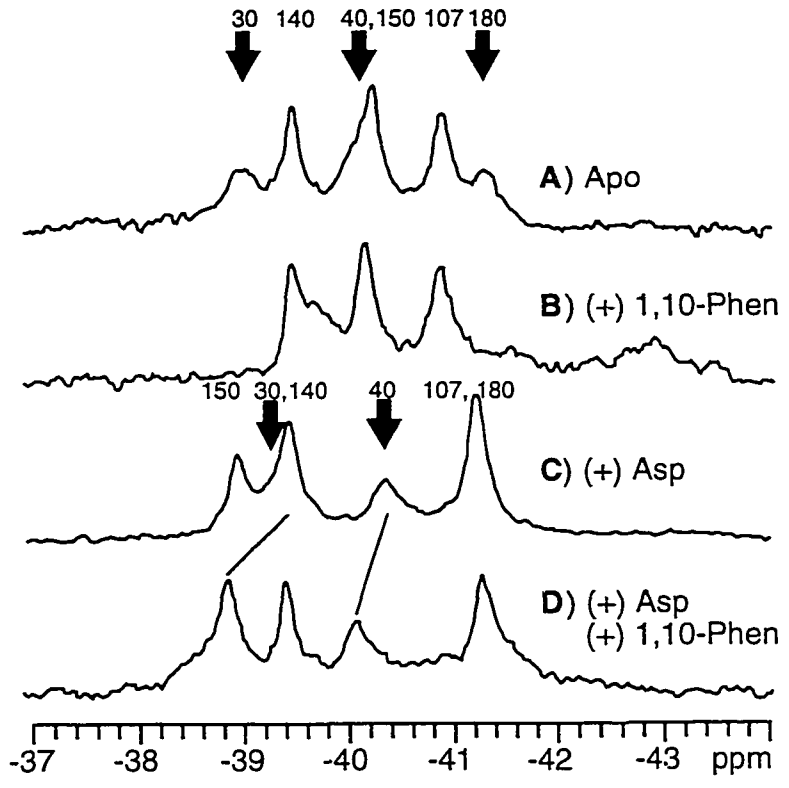


Figure 2.6. Effect of 1,10-phenanthroline on the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand binding domain. Shown are the spectra of the (**A**) apo, (**B**) phenanthroline-bound, (**C**) aspartate-bound, and aspartate and (**D**) phenanthroline-bound states of the oxidized N36C ligand binding domain. The bold arrows indicate the assigned resonances which are shifted by ligand binding; where known, the new positions of these resonances are indicated by the light diagonal lines. Sample parameters were as in Figure 2.3; also present were 3 mM 1,10-phenanthroline and 5 mM aspartate where indicated. The concentration of dimeric domain was 0.4 mM.

observed in the crystal structure (142), are significantly shifted by 1,10-phenanthroline binding. In contrast, the more distant 4-F-Phe107, 4-F-Phe140, and 4-F-Phe150 resonances show no detectable change. Similarly, when 1,10-phenanthroline is added to the aspartate-bound domain, the only resonances affected are 4-F-Phe30 and 4-F-Phe40 in the 1,10-phenanthroline binding site (Figure 2.6C,D and Table 2.3). Thus the observed effects of 1,10-phenanthroline binding are confined to its aromatic binding pocket at the subunit interface.

Stoichiometry of the Attractant-Induced Conformational Change. The two empty attractant-binding sites of the apo-domain are virtually identical due to the C_2 symmetry axis of the homodimeric structure, but crystallographic evidence suggests that only one site is occupied by aspartate at saturation (142, 225). Titration of the ^{19}F NMR spectrum of 4-F-Phe labeled domain (oxidized N36C) with aspartate yields information concerning the stoichiometry of aspartate required to trigger the maximal conformational change (Figure 2.7). No further change in the ^{19}F NMR spectrum is observed after the first aspartate binding site in the dimer becomes occupied. In particular, the 4-F-Phe150 resonance has shifted to a new frequency and attains its full magnitude when the mole ratio of [aspartate] /[receptor dimer] reaches 1.0 (Figure 2.7). It should be noted that although these results are consistent with half-of-sites binding, they do not rule out the subsequent binding of an aspartate at the second site without further conformational change.

Kinetics of Aspartate Binding. The kinetics of aspartate binding can be probed using the ^{19}F NMR titration data of Figure 2.7 and equations 2a-c, which describe the general relationship between the rate of interconversion of two conformations and the character of the NMR spectrum (210):

Slow Interconversion	$\upsilon_{ic} << \lvert\upsilon_a - \upsilon_b\rvert$	(2a)
Intermediate	$\upsilon_{ic} \sim \lvert\upsilon_a - \upsilon_b\rvert$	(2b)
Rapid	$\upsilon_{ic} >> \lvert\upsilon_a - \upsilon_b\rvert$	(2c)

If two conformations which exhibit resonance frequencies υ_a and υ_b are allowed to interconvert at the rate υ_{ic}, the resulting lineshapes will reveal which of the indicated limits applies. In the slow interconversion limit, υ_{ic} is slow relative to the frequency difference between the two conformations, thus distinct resonances are observed at υ_a and υ_b. In the intermediate interconversion limit, υ_{ic} matches the frequency difference and the resonances are significantly broadened such that they disappear. In the rapid

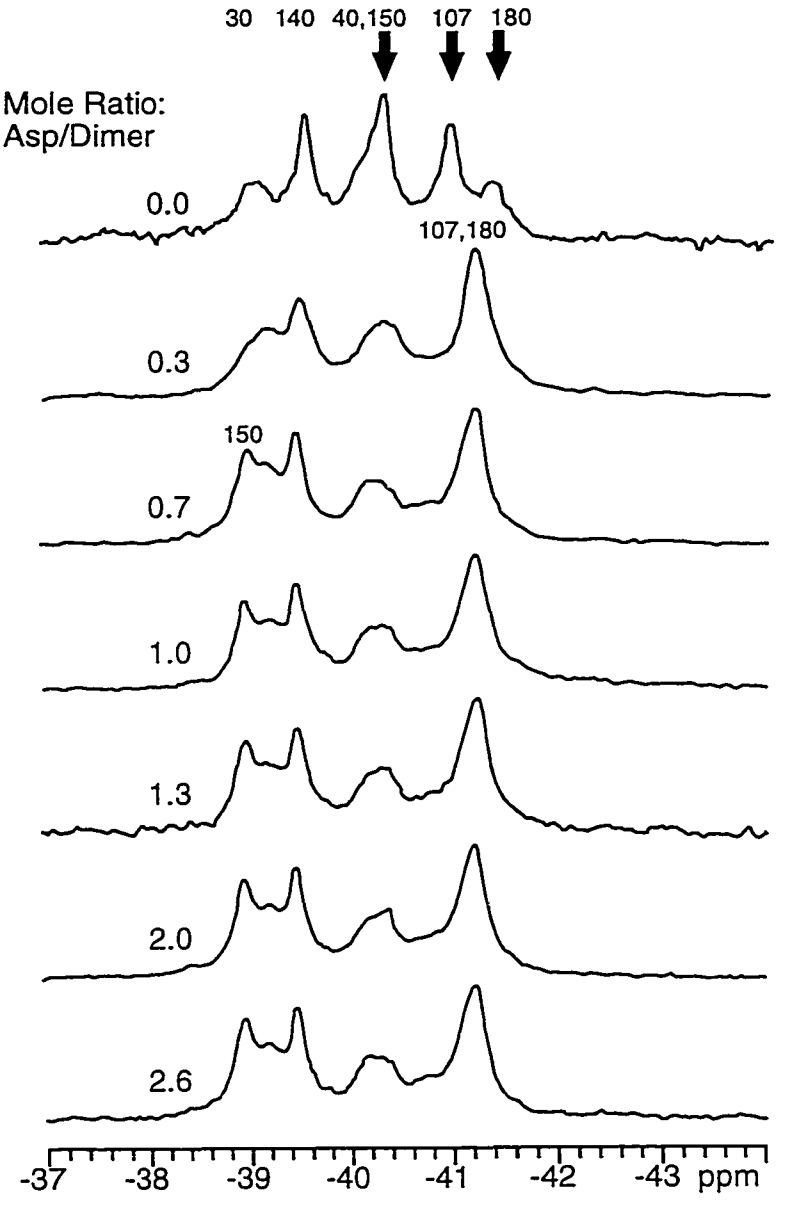


Figure 2.7. Aspartate titration of the ^{19}F NMR spectrum of the 4-F-Phe-labeled ligand binding domain. The mole ratio [total aspartate] / [total oxidized N36C dimer] is indicated for each spectrum. Bold arrows highlight the resonances shifted by aspartate binding; the final position of these resonances in the aspartate-saturated state are indicated by the revised assignments. Sample parameters were as in Figure 2.3. The free aspartate concentration in the final sample (mole ratio = 2.6) was > 1 mM, since the total aspartate concentration (4 mM) significantly exceeded the concentration of attractant binding sites (3 mM, assuming two binding sites per dimeric domain).

interconversion limit, υ_{ic} is rapid relative to the frequency difference, and a single average resonance is observed at a frequency that is the population-weighted mean of υ_a and υ_b.

The rate constant for aspartate binding was determined by titrating the domain with this ligand, while noting the ligand concentration which yielded the intermediate interconversion limit. At that point the rate of conformational interconversion caused by aspartate binding matches the frequency shift generated by the aspartate-induced conformational change. The analysis focused on the 4-F-Phe150 resonance of the oxidized N36C domain, which exhibits a frequency of -40.2 ppm in the absence of aspartate, as illustrated in Figure 2.7. When aspartate is added, the structure begins to interconvert between the apo- and aspartate-occupied conformations. As the aspartate concentration increases, the rate of aspartate binding increases accordingly, and the interconversion rate υ_{ic} passes from the slow limit into the intermediate limit. Figure 7 reveals that this intermediate limit is reached when the [aspartate] / [receptor dimer] ratio nears 0.3, where the 4-F-Phe150 resonance is observed to disappear. Further addition of aspartate continues to increase the interconversion rate υ_{ic} until it enters the rapid limit, where the 4-F-Phe150 resonance reappears at its new chemical shift of -38.9 ppm. The overall frequency shift of the 4-F-Phe150 resonance between the apo- and aspartate-saturated conformations is 1.3 ppm ~ 610 Hz, corresponding to an NMR timescale of ~ 2 msec. Using this net frequency shift and the parameters of the intermediate interconversion limit (Eqn. 2b), the rate constants for aspartate binding and dissociation are calculated as follows. In the intermediate limit the interconversion rate, or the aspartate on-rate, equals the net frequency shift of the 4-F-Phe150 resonance (or, $\upsilon_{ic} = k_{on}$ [aspartate] ~ 610 Hz). This on-rate, together with the estimated concentration of free aspartate ([aspartate] ~ 0.4 μM) in the intermediate exchange limit yield the pseudo first order rate constant for aspartate binding $k_{on} \sim 10^9\ M^{-1}sec^{-1}$. Subsequently this on-rate constant, together with the known dissociation constant ($K_D = k_{off} / k_{on} \sim 1$ μM) can be used to deduce the off-rate for aspartate dissociation, yielding $k_{off} \sim 10^3\ sec^{-1}$.

Kinetics of Subunit Resonance-Averaging. The ^{19}F NMR results indicate that the structures of the two subunits rapidly sample the same range of conformations, (Eqn. 2c) so that their time-averaged spectra are similar or identical. Thus, only six resonances are observed for the twelve 4-F-Phe residues of the oxidized N36C apo-dimer (Figure 2.4A), and again for the aspartate-saturated dimer only six resonances are detected (Figure 2.4B). It follows that in both of these ligation states the two subunits are equivalent on average. In contrast, if the two subunits had different conformations which interchanged slowly or were different on average, then up to twelve 4-F-Phe resonances would be observed for a given ligation state. For the apo-dimer, the C_2 symmetry axis proposed from the crystal structure may account for the observed equivalence of the two subunits (142). However, for the dimer containing one molecule of bound aspartate, the observed equivalence implies that the aspartate occupancy of the two binding sites must alternate rapidly. Since the frequency differences between the instantaneous resonances of the two subunits in the dimer cannot be measured, the timescale of the observed resonance-averaging cannot be calculated directly. Instead these dynamics can be inferred from the deduced aspartate dissociation rate $k_{off} \sim 10^3$ sec^{-1}, which guarantees that the aspartate occupancy of the two sites will randomly alternate on the msec timescale.

Effects of the Cys36-Cys36' Inter-Subunit Disulfide Bond. Shown in Figure 2.8 and Table 2.3 are the spectra of the wildtype domain lacking the Cys36-Cys36' disulfide, the position of which is shown in Figure 1, as well as the reduced and oxidized N36C domains, each in the presence and absence of aspartate. The Cys36-Cys36' disulfide was observed to be formed upon isolation, and was reduced by overnight incubation in the presence of 50 mM DTT. As indicated in Figure 2.8, DTT had no effect on the structure of the wildtype ligand binding domain, indicating that spectral changes caused by DTT in the N36C domain stemmed specifically from reduction of the disulfide bond.

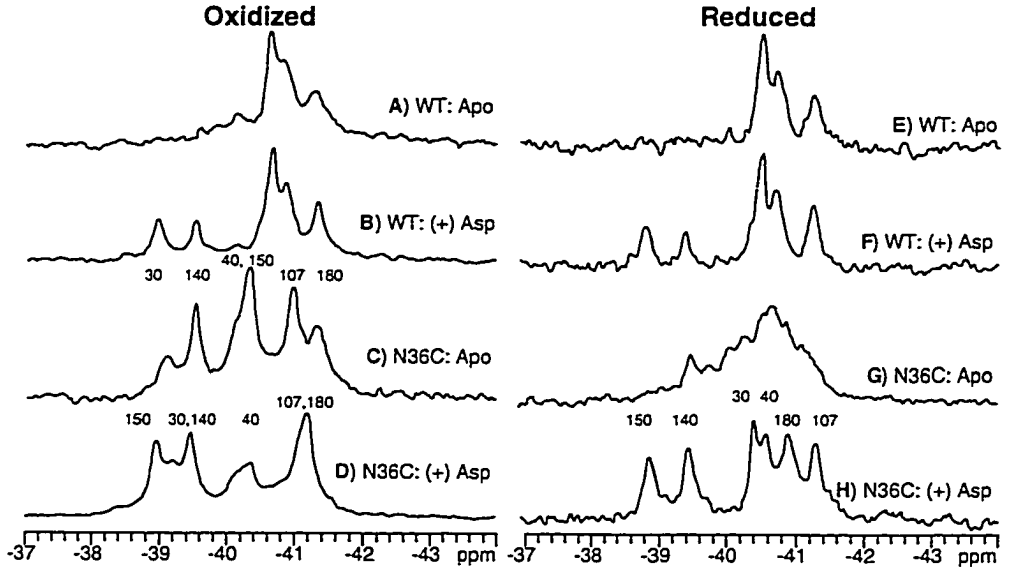


FIGURE 2.8. Effect of the Cys36-Cys36' crosslink on the ^{19}F NMR spectrum of the 4-F-Phe labeled ligand binding domain. Shown are the spectra of domains under (**A-D**) oxidizing or (**E-H**) reducing (50 mM DTT) conditions for both the wildtype and N36C mutant domains. The known assignments are indicated. Sample parameters were as in Figure 2.3; also present was 5 mM aspartate, where indicated. The concentration of dimeric domain was 0.4-1.0 mM.

When the Cys36-Cys36' disulfide is absent, either in the wildtype domain or the reduced N36C domain, the ^{19}F NMR spectrum exhibited poorly resolved resonances (Figure 2.8A,E,G), indicating a heterogeneous or dynamic structure. Addition of aspartate or formation of the Cys36-Cys36' disulfide serves to stabilize the structure, as demonstrated by the dramatic increase in the resolution of the six resonances (Figure 2.8B-D,F,H).

In wild type or reduced domains lacking the Cys36-Cys36' disulfide, the enhanced resolution observed in the aspartate-occupied state allows the effect of the disulfide on this state to be examined. Three resonances, 4-F-Phe30, 4-F-Phe40, and 4-F-Phe180, are shifted both by the presence of the N36C mutation and by oxidation of the

Cys36-Cys36' crosslink. In the crystal structure, each of these three residues are in close proximity to the Cys36 position (Figure 2.1). The remaining resonances, corresponding to residues 4-F-Phe107, 4-F-Phe140, and 4-F-Phe150 which are each >10Å from position 36, are affected by neither the Cys36 mutation nor the disulfide, as illustrated in Figure 2.8. Thus the structural changes associated with Cys36 substitution and disulfide formation do not extend to these more distant positions.

DISCUSSION

The current ^{19}F NMR titration data reveal that the binding of one aspartate molecule per homodimer triggers the maximum conformational change in the isolated ligand binding domain. This is consistent with the existence of negative cooperativity between the two attractant binding sites, as previously observed (22, 142). Positive cooperativity and non-cooperative binding are ruled out, since each would require a higher aspartate binding stoichiometry to generate the maximal conformational change. The observed negative cooperativity requires that the binding of aspartate destroys the C_2 symmetry of the ligand binding domain. Such an observation leads to a model in which the binding of an attractant molecule to the first site generates an allosteric conformational change which repositions residues in the second site such that the ligand affinity of that site is reduced or abolished. It should be noted that aspartate binds to either site in the apo-domain, since the ^{19}F NMR resonances of the two subunits in the dimer are identical, indicating that aspartate binding alternates rapidly between the two sites on the NMR timescale to yield resonance averaging. In contrast, if one of the two sites in the apo-domain was permanently incapable of binding aspartate, the ^{19}F NMR resonances from different subunits would be twinned at saturating aspartate, since the two subunits would be permanently inequivalent.

Kinetic studies of chemotaxis *in vivo* reveal that the chemotactic response time for aspartate is ~ 0.2 sec (101, 180). The ^{19}F NMR results demonstrate that the

conformational activation of the ligand binding domain is extremely rapid on this timescale; the estimated on-rate constant for aspartate binding is $k_{on} \sim 10^9$ $sec^{-1}M^{-1}$. This on-rate constant is 1-3 orders of magnitude slower than diffusion-controlled binding (64), thus binding is somewhat slowed by steric factors or the accompanying conformational change. Yet by the standards of protein binding sites the attractant site is relatively accessible, since the observed on-rate constant is comparable to that observed for enzymes such as malate dehydrogenase, lactate dehydrogenase, and chymotrypsin (46, 47, 168). Using the on-rate constant for aspartate binding and the known aspartate dissociation constant, the off-rate constant is estimated to be $k_{off} \sim 10^3$ sec^{-1}. In principle the conformational changes associated with attractant binding and release could be slower in the intact transmembrane receptor where motions in the transmembrane and cytoplasmic domains are coupled to attractant binding. However the aspartate K_D of the ligand binding domain is relatively unchanged when it is severed from the rest of the protein; thus it seems likely that even in the intact receptor, attractant binding, dissociation, and the coupled conformational changes will also be rapid, and will not contribute significantly to the chemotactic response time. Further evidence supporting the proposal that the intact receptor is a highly dynamic molecule is provided by its extensive flexibility, as observed in disulfide-trapping studies of thermal backbone motions (62).

The attractant-induced changes in the chemical shifts of the resonances 4-F-Phe150, 4-F-Phe107, and 4-F-Phe180 reveal a long-range conformational change within the ligand binding domain. Additional evidence is provided by the attractant-induced solvent exposure of the 4-F-Phe180 residue (Figure 2.5) and the sensitivity of the 4-F-Phe 180 resonance to 1,10-phenanthroline (Figue 2.6). Moreover, since 4-F-Phe107 is >10 Å from the subunit-subunit interface and is buried within the core of a single subunit, the change in the chemical shift of this resonance indicates that a long-range *intra-subunit* conformational change is triggered by attractant binding. By contrast, no inter-subunit conformational changes are observed for 4-F-Phe30 and 4-F-Phe40 at the subunit

interface of the disulfide-stabilized dimer. It could be argued that the disulfide bond blocks rearrangement of the subunit interface, but the Cys36-Cys36' disulfide bond is nonperturbing in the full length receptor where this disulfide retains normal kinase regulation. Thus, we now propose that primarily intra-subunit structural changes are induced by the binding of an aspartate or glutamate molecule. This long-range conformational change is specific for attractants, since the binding of the control molecule 1,10-phenanthroline yields a much different, localized conformational change that specifically perturbs 4-F-Phe30, 4-F-Phe40, and 4-F-Phe180, each of which helps to coordinate this molecule. It is likely that the attractant-induced intrasubunit conformational change is involved in transmembrane signal transduction. Studies in which the transmembrane and cytoplasmic domains of one subunit in the dimer are removed reveal that transmembrane signaling can occur within the remaining intact subunit, although the signaling is significantly reduced (70, 144, 201). Thus, the intra-subunit conformational change alone appears to be sufficient for transmembrane signaling, although it must be accompanied by a more localized inter-subunit rearrangement in the vicinity of the aspartate binding pocket to explain negative cooperativity. A schematic model summarizing the structural and kinetic aspects of aspartate binding is presented in Figure 2.9.

The Cys36-Cys36' disulfide bond leaves transmembrane signaling intact (61), and is relatively nonperturbing of both the apo- and aspartate-occupied conformers of the full length receptor. Consistent with this model is the observation by ^{19}F NMR that the disulfide bond detectably alters only the local structure in the vicinity of the Cys36 residue in the ligand binding domain. The structural dynamics or heterogeneity observed by ^{19}F NMR in the absence of the disulfide may well be an artifact of removal of the transmembrane helices: in this picture, the Cys36-Cys36' disulfide is proposed to stabilize a conformation of the ligand binding domain which more closely resembles its conformation in the intact receptor. Crystallographic studies have revealed that helices α1

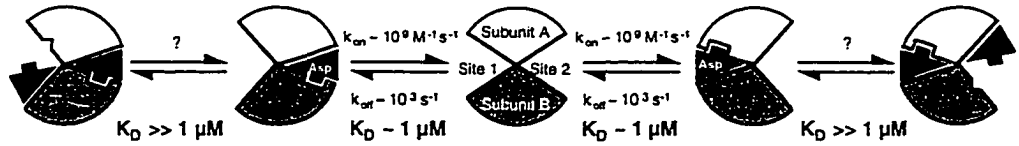


Figure 2.9. Schematic model of the structural and kinetic aspects of aspartate binding. In the absence of bound aspartate, the ligand binding domain has, on average, a symmetrical structure possessing a C_2 axis (center) lying between the two identical subunits. The binding of the first aspartate molecule can occur at either of the two equivalent attractant binding sites, with the indicated association and dissociation rate constants. This binding event causes a conformational change within at least one subunit, and destroys the symmetry of the dimer. The conformational change is communicated to the empty site, where the structure is altered such that the affinity for aspartate is significantly reduced (negative cooperativity). The binding of a second aspartate to this empty site was neither detected nor excluded by the present study, however if it occurs it generates little or no conformational change within the protein at the positions monitored by ^{19}F NMR.

and α1' of the isolated domain spread apart when the Cys36-Cys36' disulfide is absent (225), consistent with the conclusion that this disulfide stabilizes interactions between the α1 and α1' helices that are present in the transmembrane receptor but weakened in the isolated domain. This conclusion also explains the greater propensity of the isolated domain to dissociate into individual subunits (145). Thus the enhanced disorder observed in ^{19}F NMR spectra when the inter-subunit disulfide is removed may well stem from the loss of native interactions between helices α1 and α1'.

Chapter Three

Cysteine and Disulfide Scanning Reveals a Regulatory α-Helix in the Cytoplasmic Domain of the Aspartate Receptor

ABSTRACT

The transmembrane, homodimeric aspartate receptor of *E. coli* and *S. typhimurium* controls the chemotactic response to aspartate, an attractant, by regulating the activity of a cytoplasmic histidine kinase. The cytoplasmic domain of the receptor plays a central role in both kinase regulation and sensory adaptation, although its structure and regulatory mechanisms are unknown. The present study, a collaboration with Randal B. Bass, utilizes cysteine and disulfide scanning to probe residues L250 through Q309, a region which contains the first of two adaptative methylation segments within the cytoplasmic domain. Following the introduction of consecutive cysteine residues by scanning mutagenesis, the measurement of sulfhydryl chemical reactivities reveals an α-helical pattern of exposed and buried positions spanning residues 270 through 309. This detected helix, termed the "first methylation helix," is strongly amphiphilic: its exposed face is highly anionic and possesses three methylation sites, while its buried face is hydrophobic. *In vivo* and *in vitro* assays of receptor function (the latter carried out by R. Bass) indicate that inhibitory cysteine substitutions are most prevalent on the buried face of the first methylation helix, demonstrating that this face is involved in a critical packing interaction. The buried face is further analyzed by disulfide scanning, which reveals three "lock-on" disulfides that covalently trap the receptor in its kinase-activating state (R.

Bass). Each of the lock-on disulfides crosslinks the buried faces of the two symmetric first methylation helices of the dimer, placing these helices in direct contact at the subunit interface. Comparative sequence analysis of 56 related receptors suggests that the identified helix is a conserved feature of this large receptor family, wherein it is likely to play a general role in adaptation and kinase regulation. Interestingly, the rapid rates and promiscuous nature of disulfide formation reactions within the scanned region reveal that the cytoplasmic domain of the full-length, membrane-bound receptor has a highly dynamic structure. Overall, the results demonstrate that cysteine and disulfide scanning can identify secondary structure elements and functionally important packing interfaces, even in proteins that are inaccessible to other structural methods.

INTRODUCTION

The aspartate receptor of *E. coli* and *S. typhimurium* is representative of a large family of cell-surface receptors that regulate two-component signaling pathways, which are widespread in prokaryotic and eukaryotic organisms (4, 23, 59, 125, 160, 191, 193, 198, 221). These receptors contain two putative transmembrane helices per subunit and, in all cases tested, form stable homodimers that signal via a transmembrane conformational change. Chimeric receptors containing the sensory domain of the aspartate receptor and the regulatory domain of another family member are functional, suggesting that members of this receptor family use a conserved mechanism of transmembrane signaling to regulate cytoplasmic histidine kinase activity (18, 109, 206). More generally, conformational transmembrane signals may provide an important component of kinase regulation in other, unrelated receptor families as well. For example, a chimera between the aspartate receptor and the human insulin receptor has been shown to possess aspartate-regulated tyrosine kinase activity, a result which has been interpreted to indicate that the cytoplasmic domain of the insulin receptor is regulated by an intradimer conformational signal (21, 147).

The aspartate receptor, like the other transmembrane receptors of the bacterial chemotaxis pathway, associates with soluble components of the pathway to form a supermolecular signaling assembly. The core of this signaling assembly is a ternary complex involving the receptor and two cytoplasmic proteins: the histidine kinase CheA and the coupling protein CheW. Once formed, the ternary complex is kinetically stable for tens of minutes, and the stability is relatively independent of the receptor ligation state (71, 177, 212). The apo receptor stimulates the autophosphorylation activity of the kinase, while attractant binding downregulates the kinase over 10^2-fold (25, 153). Other cytoplasmic proteins associate transiently with the ternary complex, including the adaptation enzyme CheR, which binds to the C-terminal tail of the receptor (219) and serves to methylate specific receptor glutamate residues (187, 202). Another adaptation enzyme, CheB, hydrolytically demethylates these regulatory side chains, and the ternary complex stimulates this demethylation activity by phospho-activating the CheB protein (86, 99, 173). The resulting feedback adaptation loop enables the ternary complex to maintain its output at an optimal level, even in the presence of a constant background stimulus of chemoattractant or repellant. Ultimately, the function of the ternary complex is to control the swimming state of the flagellar motor by phosphorylating the CheY protein (16, 25). Phospho-CheY subsequently dissociates from the ternary complex and diffuses to the flagellar motor, where it binds and alters the swimming behavior of the cell (for reviews of the chemotaxis pathway, see 23, 59, 60, 135, 160, 191, 193).

The aspartate receptor is a 120 KDa homodimer both in the presence and absence of aspartate (41, 62, 143). The structures of the periplasmic and transmembrane domains have been extensively characterized. The 1.85 Å resolution crystal structure of the isolated periplasmic domain reveals a dimer of four-helix bundles (α1-α4 and α1'-α4', respectively) (28, 142, 226). Two symmetric attractant binding sites lie at the dimer interface where they are coupled to yield negative cooperativity, such that occupancy of one site reduces or prevents occupancy of the second site (22, 50, 142, 226). The N- and

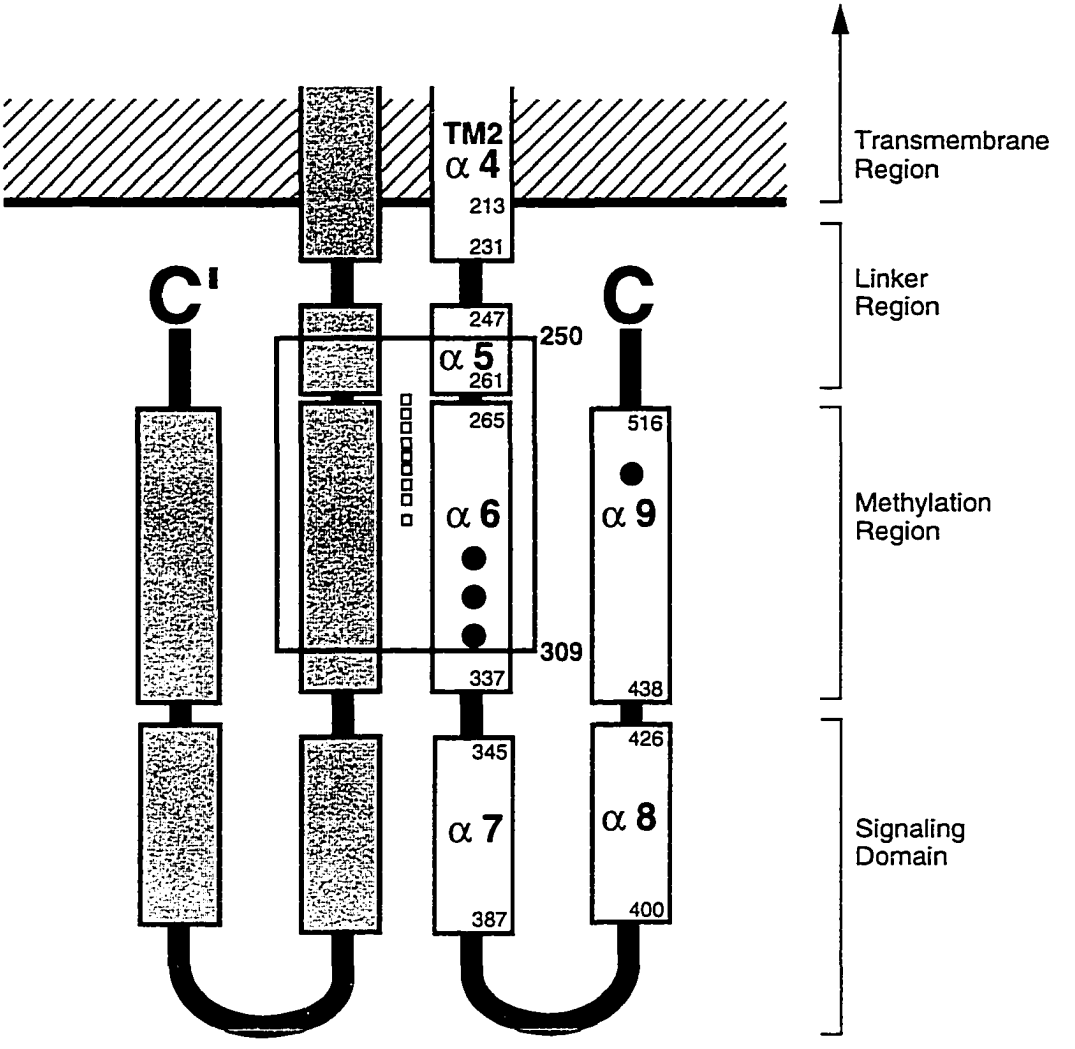


Figure 3.1. Schematic secondary structure model for the homodimeric cytoplasmic domain of the aspartate receptor based on the pattern recognition in a sequence alignment (Chapter 4; 114). Rectangles indicate the regions predicted to be α-helical, and the sites of regulatory methylation are indicated by the filled circles. The small open squares indicate the sites of point mutations which restore signaling function to a defective receptor possesing the A19K mutation in its first transmembrane helix (156). The current study utilizes cysteine and disulfide scanning to probe the region highlighted by the large box (L250 through Q309). Although only one subunit is shown in detail, the two identical subunits are related by a two-fold rotational symmetry axis.

C-terminal helices of each periplasmic subunit extend across the bilayer, yielding a membrane-spanning four-helix bundle. In both the periplasmic and transmembrane regions, the subunit interface is dominated by the packing interactions between the symmetric first transmembrane helices (α1/TM1 and α1'/TM1') along the central axis of the dimer, while the second transmembrane helices (α4/TM2 and α4'/TM2') pack along the periphery of the central pair (28, 39, 41, 62, 137, 142, 158). Previous studies have shown that the second transmembrane helix is the signaling helix that carries the ligand-induced signal across the bilayer (39, 50). Attractant binding to the receptor generates a piston (or swinging-piston) displacement of the signaling helix toward the cytoplasm (40). Parallel studies have shown that the related chemoreceptor for ribose and galactose shares the same helical structure and piston-type signaling mechanism, illustrating the generality of this transmembrane signaling motif at least within the bacterial chemoreceptor subfamily of two-transmembrane helix receptors (17, 90, 91, 115-117).

In contrast to the periplasmic and transmembrane domains, the structure and mechanism of the cytoplasmic domain remains poorly understood. It has been demonstrated that the isolated cytoplasmic domain fragment, though functional, is highly dynamic and exhibits heterogeneous oligomeric states (124, 149, 179, 195). These features have hindered attempts to determine the structure of the domain by crystallography and NMR. Thus, other structural methods are needed. Circular dichroism and hydrodynamic measurements have implied that the domain is an elongated bundle of α-helices (124, 149, 220), while comparative sequence analysis also suggests a predominantly helical structure. The latter analysis, which has yielded the predicted helical arrangement summarized in Figure 3.1, is based on an alignment of 56 cytoplasmic domain sequences from chemo-, thermo- and photo-receptors (Chapter 4; 114, 152, 188, 192, 224). The current study focuses on residues L250-Q309, which includes the first methylation segment of the cytoplasmic domain. This segment contains three of the four methylation sites of the receptor (202) and lies between the signaling helix α4/TM2 and

the kinase interaction domain, suggesting that it is likely to play an important role in signal transduction. Genetic studies have underscored the importance of the first methylation segment to receptor signaling. Second site repressor mutations that counteract the deactivating mutant A19K, which lies in the first transmembrane helix, are found to cluster within this region (156), as do many signal-locking point mutations in the homologous receptor Tsr (9-11). Finally, biochemical studies analyzing the isolated cytoplasmic domain have implicated the first methylation segment in kinase regulation (43, 195). The periodicity of CheR recognition residues and comparative sequence analysis of homologous receptors suggest that the targetted region contains at least one α-helix (Chapter 4; 114, 192, 202, 203), as shown in Figure 3.1. These predictions remain to be tested, however.

The current study probes the structure of the first methylation segment in the full-length, membrane-bound receptor and tests specific mechanisms of kinase regulation involving this region. The experimental approach begins with cysteine scanning mutagenesis, a systematic approach analagous to alanine scanning mutagenesis (216) used to identify secondary structure elements in proteins (2, 3, 67, 116, 175). Cysteine scanning provides several advantages over alanine scanning, however:

1) Numerous alkylating agents, including reactive groups on many fluorophores (84), react specifically with cysteine at physiologically relevant pH's (the range of effective pH centers around neutrality). The rate of reaction provides information about solvent accessibility and chemical environment.
2) Cysteine has the unique property of forming a covalent disulfide bond between pairs of cysteine residues. Thus, the rate of disulfide formation between pairs of cysteines can in most cases be easily measured and can serve as an indicator of the sulfhydryl-sulfhydryl separation, local structural dynamics, and the chemical environment (32-34, 61, 62, 158).

3) The conditions of the above two reactions can be easily optimized to provide the appropriate reaction characteristics for different proteins, or different regions of the same protein.
4) Formation of the disulfide bond imposes strict geometric constraints between the two cysteines involved (33), particularly the interatomic separation between β-carbons. Thus, when protein activity is maintained even in the presence of an intramolecular disulfide bond, the two cysteine positions are proximal in the native structure.
5) Cysteine residues occur with low natural abundance (45). Thus, engineered cysteines are often unique and are easily introduced into cloned proteins. Furthermore, native cysteines can often be removed without functional consequence (167, 207, J.B. Kanopka, personal communication), and in cases where they are functionally important they are sometimes inaccessible to the chemistry mentioned above, due to extensive burial or disulfide formation during protein folding (Bass & Falke, unpublished). Thus, the utility of cysteine substitution is not necessarily limited to proteins naturally lacking this residue.
6) Though chemical crosslinkers are not used in this thesis, a variety of bifunctional reagents are available which connect cysteine pairs within a variety of interatomic distance ranges (218). Thus, a range of distance constraints can be placed between cysteine pairs while testing the protein's functionality.

Analysis of the chemical reactivities and functional effects of the cysteines substituted at positions 250 through 309 reveals considerable α-helical structure in this region and also maps out a functionally critical buried face of an amphiphilic helix. Moreover, the disulfide scanning method (39, 41, 62, 115, 158) confirms the previously

observed dynamics of the cytoplasmic domain (179) and is used to discover inter-subunit disulfide bonds that lock the kinase in its on-state. These lock-on disulfides identify a specific helix-helix contact that exists in the on-state of the receptor. Overall, the results provide the first experimental characterization of helical secondary structure and packing within cytoplasmic domain, and show that a helix-helix interface between receptor subunits plays an important role in the mechanism of kinase regulation.

MATERIALS & METHODS

Materials. The aspartate receptor was expressed in *E. coli* RP3808 (*Δ(cheA-cheZ)DE2209 tsr-1 leuB6 his-4 eda-50 rpsL136 [thi-1 Δ(gal-attl)DE99 ara-14 lacY1 mtl-1 xyl-5 tonA31 tsx-78]/mks/*) for isolation of receptor-containing membranes *in vitro*, and in *E. coli* RP8611 (*Δ(tsr)DE7028 Δ(tar-tap)DE5201 zbd::Tn5 Δ(trg)DE100 leuB6 his-4 rpsL136 [thi-1 ara-14 lacY1 mtl-1 xyl-5 tonA31 tsx-78] /CP362 of G.Hazelbauer via F. Dahlquist, pa/*) for analysis of swarm rates *in vivo*. Both strains were kindly provided by John S. Parkinson (University of Utah, 123). Strains and plasmids used for expression of the histidine kinase CheA (HB101/pMO4) and the coupling protein CheW (HB101/pME5), were kindly provided by Jeff Stock (Princeton University). The strain and plasmid used for expression of CheY (RBB455/pRBB40) were graciously provided by Bob Bourret (University of North Carolina). 5-Iodoacetamido-fluorescein was purchased from Molecular Probes, Inc. [γ-^{32}P]ATP (6000 Ci/mmol) was obtained from Amersham. Mutagenic oligonucleotides were synthesized by Gibco-BRL.

Cloning and Mutagenesis. The gene encoding the *S. typhimurium* aspartate receptor under control of its natural promotor was cloned into the vector pBLUESCRIPT KS$^-$ (Stratagene), yielding the plasmid pSCF6 as previously described (41). This plasmid was used for both mutagenesis and overexpression of wildtype and mutant receptors. Site-directed mutagenesis was carried out according to Kunkel et al. (111), with the modifications described by the mutagene phagemid mutagenesis kit (BioRad).

The resulting plasmids were transformed into *E. coli* LM1035 (a derivative of HB101), then isolated using QIAprep spin columns (QIAGEN). The sequences of the mutants were confirmed by plasmid DNA sequencing, performed by thermocycling with Sequitherm DNA polymerase (Epicentre Technologies).

Preparation of Membranes Containing the Aspartate Receptor. Wildtype and mutant receptors were expressed in *E. coli* RP3808 bearing the appropriate version of pSCF6. Saturated liquid cultures were grown at 37°C with shaking in Luria Broth (15). These 2 ml cultures were then diluted 1/250 into a Vogel Bonner Citrate minimal growth medium (209), containing 0.75% glycerol, 200 μg/ml $MgSO_4 \cdot 7H_2O$, 2000 μg/ml citric acid$\cdot H_2O$, 10,000 μg/ml K_2HPO_4, 3500 μg/ml $NaNH_4 \cdot HPO_4 \cdot 4H_2O$, 40 μg/ml D,L-histidine, 20 μg/ml L-methionine,20 μg/ml L-leucine,20 μg/ml L-threonine,1 μg/ml thiamine, and 100 μg/ml ampicillin (65). Cells were incubated with vigerous shaking at 30°C for 18 h, then harvested by centrifugation [Beckman JA-10 rotor at 8,000 RPM (11,300 x g) for 10 min]. A 500 ml prep typically yielded from 1.5 to 2.0 g of cells.

Receptor-containing bacterial membranes were isolated using a variation of a procedure previously described by Chervitz & Falke (39). The cell pellets were resuspended in 8 ml of a low salt buffer containing 100 mM sodium phosphate, pH 7.0 with NaOH, 10% glycerol, and 10 mM EDTA, 50 mM DTT, 0.5 mM PMSF, and 2.5 mM 1,10-phenanthroline. The cells were then lysed by sonication with a macro tip (Mysonix, Inc.) at 70% maximum power five times for 20 s each, with 1 min intervals between sonications. Sonication was performed in an ice-NaCl-water bath using thin-walled plastic centrifuge tubes to facilitate heat dissipation (1 x 3.5 inches, Beckman Ultra-Clear tubes). After sonication, unbroken cells and particulate matter were removed by ultra-centrifugation [Beckman TLA100.3 rotor at 15,000 RPM (12,000 x g) for 20 min]. Membranes were then pelleted by ultra-centrifugation [Beckman TLA100.3 rotor at 100,000 RPM (540,000 x g) for 15 min]. The membranes were resuspended in 3 ml of a

high salt buffer containing 20 mM sodium phosphate, pH 7.0 with NaOH, 2M KCl, 10% glycerol, 10 mM EDTA, 5 mM DTT, 0.5 mM PMSF, and 2.5 mM 1,10-phenanthroline. Resuspension was achieved by sonication three times for 20 s each using a 1/8 inch microtip set at 40% maximum power, with 1 min intervals between sonications. Membranes were pelleted as before. The high salt wash was then repeated as above except that the buffer contained no DTT or 1,10-phenanthroline. Finally, the membranes were washed as before in 3 ml of the final buffer, containing 20 mM sodium phosphate, pH 7.0 with NaOH, 10% glycerol, 0.1 mM EDTA, and 0.5 mM PMSF. The membrane pellets were then resuspended in 300 µl of fresh buffer by sonication with the microtip at 10% maximum power. The final suspensions were aliquoted, frozen in liquid nitrogen, and stored at -70°C.

The total protein yield was determined by a BCA assay (194) in which the color-developing incubation was carried out at 65°C in the presence of 0.1% SDS. A typical 500 ml prep yielded ~20 mg of total protein. To ascertain the fraction of this total protein comprised by the receptor, the membrane components were resolved on 10% SDS-PAGE gel with an acrylamide:bis-acrylamide ratio of 40:0.2. The relative intensities of the coomassie-stained protein bands were quantitated by laser densitometry (Ultroscan XL, LKB). The receptor typically comprised from 10 to 20% of the total protein concentration.

Preparation of Soluble Chemotaxis Components. CheY was purified as previously described (27). CheA (189) and CheW (190) were purified using an adaptation of the published protocols. Saturated liquid cultures of the appropriate strain/plasmid combination were grown in Tryptone Broth (10 g/l tryptone, 4 g/l NaCl) overnight at 37°C with vigorous shaking. These cultures were diluted 1/500 into Tryptone Broth and grown at 37°C. At mid-log phase (OD_{600} ~ 0.8, approximately 6-8 h) expression was induced by adding 3-β-indole acrylic acid to 0.1 mg/ml. The cultures

were then incubated for an additional 8-10 h before the cells were harvested by centrifugation [Beckman JA-10 rotor at 6,000 RPM (6370 x g) for 10 min].

The isolated cells were resuspended in 20 ml of ice-cold TEDG 20 (50 mM Tris, pH 7.5 with HCl, 1 mM EDTA, 2 mM DTT, 20% glycerol) containing 1 mM PMSF, then the cells were lysed by french press and membranes and particulate matter were removed by ultra-centrifugation [Beckman TLA100.3 rotor at 80,000 RPM (340,000 x g) for 20 min]. To the supernatant, solid ammonium sulfate was added to 45% saturation. This solution was stirred on ice for 30-60 min, then the precipitated protein was pelleted by centrifugation [Beckman JA-20 rotor at 13,000 RPM (20,400 x g) for 15 min]. The pellet was resuspended in 4 ml of TEDG 20 with 50 mM NaCl and 1 mM PMSF, and dialyzed overnight against 500 ml of the same buffer, with one change of buffer. CheA or CheW was isolated from the dialysate by FPLC on a Q-Sepharose column (Pharmacia) with a gradient of 50 mM to 1 M NaCl in TEDG (50 mM Tris, pH 7.5 with HCl, 1 mM EDTA, 2 mM DTT, 10% glycerol). The final column step was FPLC on Superdex 200 16/60 column (Pharmacia) for CheA or a Superdex 75 16/60 (Pharmacia) column for CheW, eluting with 20 mM Tris, pH 7.5 with HCl, 0.5 mM EDTA, 0.5 mM DTT, 10% glycerol, 750 mM NaCl. The pooled fractions from each column were concentrated by ultrafiltration (Amicon YM10 membrane) to a final volume of 1 to 2 ml yielding a sample that was dialyzed overnight against 500 ml of TEDG with one change of buffer. Particulate matter was removed from the dialysate by ultra-centrifugation [Beckman TLA100.3 rotor at 100,000 RPM (540,000 x g) for 10 min]. The sample was then aliquoted and snap-frozen in liquid nitrogen before storage at -70°C.

The total protein yield was determined by a BCA assay as described above. To ascertain the purity of the protein, the sample was resolved on a 10% (CheA) or 15% (CheW) acrylamide SDS-PAGE gel, quantitating by laser densitometry. A typical 4 l prep yielded ~2 mg of CheA or CheW at >95% homogeneity.

Analysis of Solvent Exposure. The procedure is illustrated schematically in Figure 3.2. Receptor-containing membrane samples were diluted to give 5 μM receptor monomer in 10 mM sodium phosphate, pH 6.5 with HCl, 50 mM NaCl, 50 mM KCl, 1 mM EDTA. Reactions were initiated by adding 5-iodoacetamido-fluorescein (IAF) from a 10 mM stock in N,N'-dimethyl-formamide to a final concentration of 500 μM. Each reaction was incubated at 25°C for 10 min then divided into two aliquots. One aliquot was quenched by addition of β-mercaptoethanol to 60 mM in order to destroy the remaining IAF, while the second aliquot was left unquenched. Subsequently, 4x Laemmli non-reducing sample buffer (250 mM Tris, pH 6.8 with HCl, 4% SDS, 40% glycerol) was added to both samples to denature the protein, followed by incubation at 95°C for 3 min. Neither a longer incubation nor subsequent addition of IAF drove the unquenched aliquot to a higher extent of labeling, indicating that the labeling reaction in this aliquot had reached completion. Finally, β-mercaptoethanol was added to 60 mM in the unquenched aliquot to destroy the unreacted IAF (and to improve resolution by SDS-PAGE for unknown reasons). Both samples were resolved on a 10% SDS-PAGE gel with an acrylamide:bis-acrylamide ratio of 40:0.2. The unstained gel was photographed on a 302 nm transilluminator using a 470 nm long-pass filter and Polaroid 665 film (1 min exposure, F-stop 5.6) then Coomassie-stained. The fluorescent receptor band from the negative was quantitated by laser densitometry then normalized to the amount of receptor measured by laser densitometry of the Coomassie-stained gel. The normalized fluorescence of the sample from the 10 min, 25°C incubation was then divided by the normalized fluorescence of the denatured reaction, yielding the ratio termed "chemical reactivity." As a negative control, the experiment was carried out with wild type receptor, for which no significant fluorescent labeling was observed. Moreover, no competing reactions were observed under the conditions employed, since control reactions in which cysteine-containing receptors were denatured either immediately or after the standard 10 min incubation yielded the same extent of labeling. In particular, disulfide formation

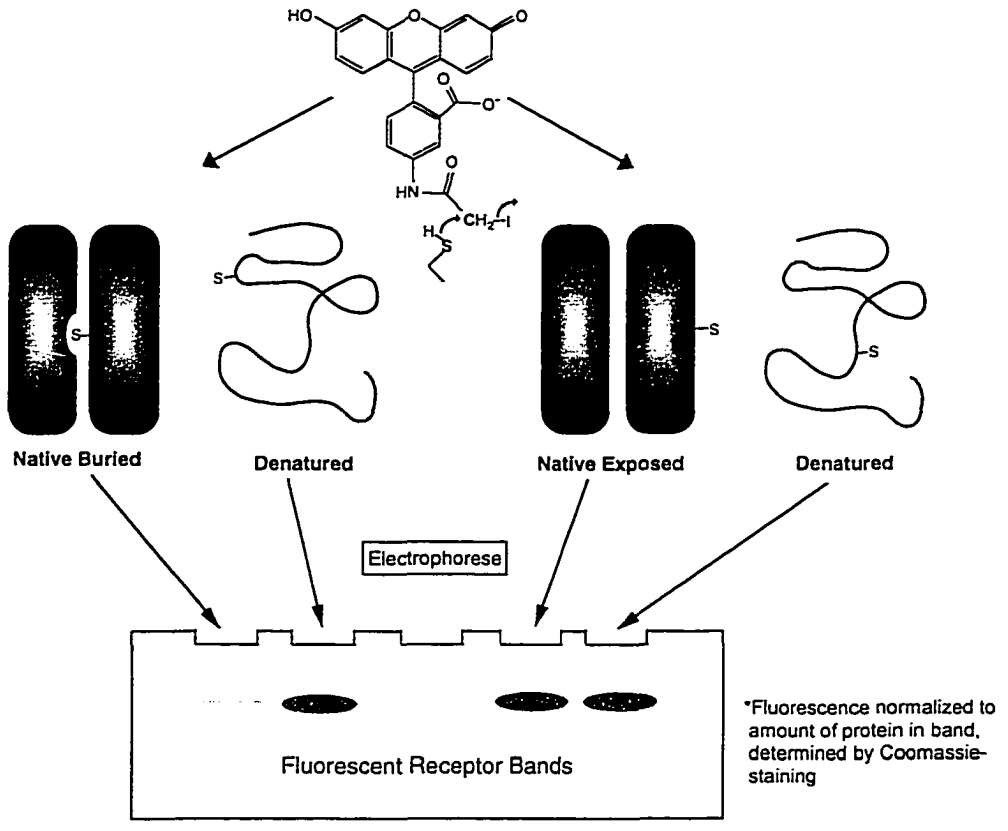


Figure 3.2. Measuring solvent accessibility through thiol reactivity. Engineered thiol-containing receptors are incubated with the large, water-soluble, thiol-specific alkylating reagent 5-iodoacetamidofluorescein (top) for a fixed interval. The reaction is then split into two aliquots. The first is quenched immediately while the second is first denatured, such that the thiol becomes fully alkylated. After resolving the labeled receptor by electrophoresis, the amount of label incorporated is determined by quantitating the fluorescence of the appropriate band, and this amount of label is normalized to the amount of receptor in the band as determined by Coomassie staining. Buried residues show a low extent of labeling in the folded receptor compared to that observed in the denatured aliquot, while exposed residues have similar extents of labeling in the two aliquots.

catalyzed by the presence of contaminating metals (sometimes termed "spontaneous" disulfide formation) was prevented by the presence of 1 mM EDTA.

***In vivo* Analysis of Receptor Function.** Chemotaxis swarm plate assays (1, 41) were carried out using *E.coli* RP8611, which lacks the aspartate receptor, transformed with the appropriate version of the plasmid pSCF6. Because the receptor is

overexpressed by the plasmid, and due to the ability of the adaptation pathway to correct subtle receptor defects, this *in vivo* assay is best suited for the detection of radical perturbations that greatly reduce or destroy receptor function. Saturated liquid cultures were grown in Luria Broth at 37°C with shaking overnight, then 5 µl was spotted onto 0.23% agar minimal plates containing Vogel Bonner Citrate medium (209) supplemented with 0.1% glycerol, 20 mM lactate, 40 µg/ml D,L-histidine, 20 µg/ml L-leucine, 1 µg/ml thiamine, and 100 µg/ml ampicillin, with or without 0.1 mM L-aspartate (215). Colony diameters were measured at 3-4 h intervals, starting approximately 18 h after spotting onto swarm plates incubated at 30°C. Swarm rates were determined by least-squares linear best-fit. To account for pseudotaxis (12) and other non-aspartate specific swarming, the rate of swarming on the plate containing no aspartate was subtracted from the rate of swarming on the plate containing aspartate. The resulting aspartate-specific swarm rate for each engineered receptor was normalized to the corresponding rate for the wildtype receptor (typically 0.6 mm/h), which was determined in parallel as a positive control. By contrast, cells transformed with vector lacking the receptor gene exhibited at least 10-fold slower expansion of the colony diameter, and for these cells, no aspartate-specific swarming was observed.

***In vitro* Analysis of Receptor-Mediated Kinase Regulation.** The ability of the receptor to regulate the activity of the CheA kinase was quantified using a coupled phosphorylation assay, carried out by R. Bass, which monitors the the production of phospho-CheY by the receptor-CheA-CheW ternary complex (25, 153). The procedure used is described in detail elsewhere (41). Briefly, isolated *E.coli* membranes containing 6 µM receptor monomer in 50 mM Tris, pH 7.5 with HCl, 50 mM KCl, and 5 mM $MgCl_2$ was combined with the purified proteins CheW (2 µM), CheA (0.25 µM monomer), and CheY (10 µM). The components were preincubated at 23°C for 30 min to permit formation of the receptor-kinase complex (71), then the reaction was initiated by the addition of [γ^{32}P]ATP (4000-8000 cpm/pmol) to a final concentration of 0.1 mM.

Aliquots (5 µl) were quenched at 10 and 30 s by mixing with 15 µL of 2x Laemmli nonreducing sample buffer. The quenched reactions were resolved on a Laemmli discontinuous gel consisting of a stacking gel of 9.6% acrylamide and 0.048% bis-acrylamide and a separatory gel of 16% acrylamide and 0.5% bisacrylamide, as well as 22% urea. The gels were dried down immediately after electrophoresis and the amount of phospho-CheY was quantitated by phosphorimaging. The initial phospho-CheY rate was determined from the slope between the 0 s and 10 s timepoints. This rate was normalized to a fixed receptor concentration (6 µM monomer) as determined by laser densitometric analysis of Coomassie-stained gels. The resulting rates for engineered receptors were divided by the corresponding rate for the wildtype receptor, measured in parallel as a positive control, yielding a phosphotransfer rate relative to the wildtype receptor-kinase complex. When the kinase was complexed with wildtype receptor, phospho-CheY was produced at a rate of 0.1 pmol s^{-1} in a reaction containing 120 pmol receptor, 40 pmol CheW, 5 pmol CheA, 200 pmol total CheY, and 2000 pmol ATP.

Oxidation of Cysteine-Containing Receptors. Formation of disulfide-linked receptors was accomplished by treating the receptor with 0.2 mM Cu^{2+}•(1,10-phenantholine)$_3$ and ambient O_2 (~250 µM) for 20 min at 37°C (33, 62). The reaction was inactivated by addition of sodium persulfate to 0.1 mM.

Protein Graphics. Crystallographic coordinates of the apo- and aspartate-occupied conformations of the ligand binding domain were graciously supplied by Kim and co-workers (142) and were visualized using Biosym Technologies Insight II graphics software running on a Silicon Graphics Personal Iris workstation. Surface accessibility calculations were carried out using the program Accessibility developed by Handschumacher and Richards (170).

Error Determination. The error ranges shown represent the standard deviation of the mean for $n > 3$.

RESULTS

Construction of Cysteine-Containing Receptors. To carry out cysteine scanning in the targetted regions of the aspartate receptor, a series of single-cysteine substitutions were generated by oligonucleotide-directed mutagenesis of the receptor gene in the plasmid pSCF6. Each substitution yielded a unique cysteine side chain, since the native receptor contains no intrinsic cysteine residues. In the periplasmic domain, a set of nine positive control receptors were generated by scanning the cysteine through positions T95 to H103 in the known structure of helix α2 (mutants generated by M. Danielson). In the cytoplasmic domain, the single cysteine residue was scanned through sixty consecutive positions from L250 to Q309, a region of unknown structure which includes the first methylation segment. (Mutants L250C through L289C were generated by R. Bass, while mutants S290C through Q309C were generated by M. Danielson).

Each engineered receptor protein was overexpressed in *E. coli* and isolated using standard procedures. The expression strain chosen, RP3808, lacks chemoreceptors and the adaptation enzymes CheR and CheB (123). As a result, membranes isolated from this strain possessed no chemoreceptors besides the engineered receptor of interest, and the adaptation state of the isolated receptor population was well-defined and homogenous. In general, the mono-cysteine receptors were expressed at levels from 25 to 100% that observed for wildtype. The only exceptions were the V265C and T266C receptors, for which no expressed protein was detectable. The adjacent nature of the latter positions suggests that they lie within a localized region that is essential for proper receptor folding or stability *in vivo*.

Determination of Solvent Exposure: Experimental Strategy. Under appropriate experimental conditions, the chemical reactivities of the cysteine scanning residues provide information regarding solvent exposure and burial, as illustrated by previous studies of transmembrane receptors, transporters and channels (2, 3, 61). In the

present study, the chemical reactivity of each engineered sulfhydryl was determined by quantifying its reaction with 5-iodoacetamido-fluorescein (IAF), a large, aqueous, sulfhydryl-specific alkylating agent. To measure this reactivity, isolated *E. coli* membranes containing the full-length receptor of interest (~5 μM total monomer) were incubated with excess fluorescent probe (500 μM) for a fixed interval (10 min at 25°C). Subsequently, one aliquot was removed and immediately quenched with β-mercaptoethanol to destroy the remaining IAF, while a second aliquot was denatured with SDS to allow the labeling reaction to proceed to completion in the unfolded receptor before quenching. The wildtype, cysteine-less receptor yielded no detectable labeling under these conditions. The cysteine-containing engineered receptors, by contrast, yielded substantial but variable levels of labeling depending on the cysteine location. Solvent-exposed cysteines were identified as those exhibiting similar extents of alkylation in the folded and SDS-denatured receptor states, while buried cysteines were recognized as those that were protected from alkylation in the folded state. The latter buried residues yielded as much as 10^3-fold lower alkylation in the folded membrane-bound state than in the denatured state.

A chemical reactivity parameter was defined as the ratio of IAF labeling in the folded and unfolded receptor states. This parameter ranges from zero (for a cysteine fully protected from labeling in the folded state) to unity (for a fully solvent-exposed cysteine). The reaction conditions were optimized to provide a large dynamic range between the measured chemical reactivities while maintaining the average value below 0.5, thereby ensuring that most reactions did not approach completion in the unfolded receptor. The method can be adapted for other regions of the same or different proteins by adjusting the reaction conditions appropriately: in general, studies of buried, static structural elements will require stronger reaction conditions, while more exposed or dynamic elements will entail milder conditions. The stringency of the labeling reaction can be controlled through

the initial concentration of the label, the pH of the reaction mixture, or the time or temperature of incubation.

The ability of the chemical reactivity parameter to resolve solvent-exposed and buried positions was tested in a control surface helix of known structure, specifically the cysteine scanning positions 95 through 103 in helix α2 of the periplasmic domain. Figure 3.3 presents the relative chemical reactivity of each engineered cysteine, and also shows the calculated solvent accessibility of the β-carbon at the corresponding position in the periplasmic domain crystal structure (142, 170). The graph displays a striking correlation between the measured chemical reactivity and the calculated solvent exposure, indicating that the chemical reactivity is determined primarily by accessibility to the large, solvent-

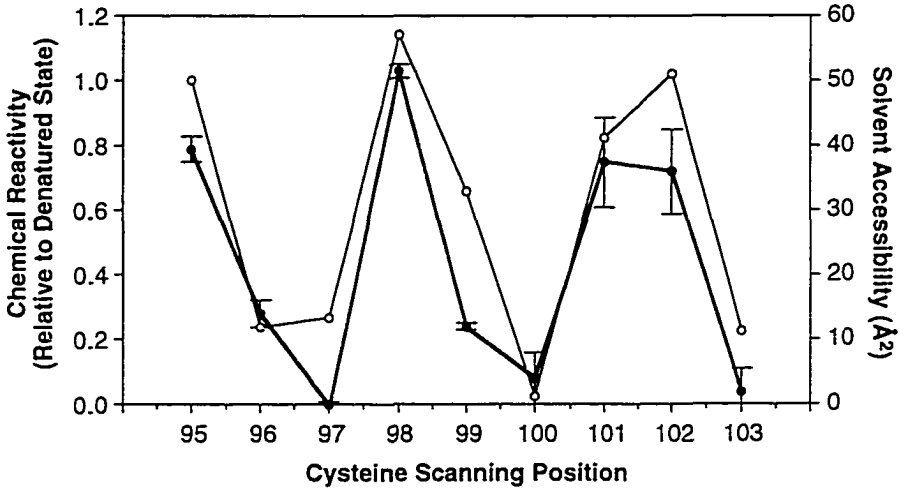


Figure 3.3. Cysteine scanning analysis of chemical reactivity and solvent exposure in the known helix α2 of the periplasmic domain. Shown is the correlation between the chemical reactivities of cysteine sulfhydryl groups and their solvent exposures determined from the crystal structure. Engineered receptors (5 μM monomer) were labelled with 500 μM 5-iodoacetamido-fluorescein for 10 min in 10 mM sodium phosphate, pH 6.5 with HCl, 50 mM NaCl, 50 mM KCl, 1 mM EDTA. While one aliquot was quenched immediately, another was first denatured with SDS to generate the unfolded, fully labelled receptor. The measured chemical reactivity (closed symbols, bold line) was defined as the ratio of the labeling observed for the folded receptor to that of the denatured receptor. Solvent exposure (open symbol, fine line) was calculated by the method of Richards (170, version 1983) using the crystal structure of the apo periplasmic domain (142).

based probe. In principle, deviation from a perfect correlation could arise from (i) a minor deviation of the crystal structure from the structure of the native protein; (ii) local electrostatic or steric forces that, in addition to the solvent exposure, modulate the alkylation rate; or (iii) a structural perturbation triggered by a given cysteine substitution. Overall, the results indicate that such deviations are not typical and that the chemical reactivity can be used to accurately map out exposed and buried positions in a representative protein structure, even when the protein is embedded in a native membrane containing a mixture of protein components.

Determination of Solvent Exposure for Targetted Cytoplasmic Positions. Cysteines scanned through positions 250 to 309 of the cytoplasmic domain reveal an α-helical pattern of chemical reactivities, as displayed in the cysteine scanning plot of Figure 3.4. Highly exposed positions are defined as those exhibiting chemical

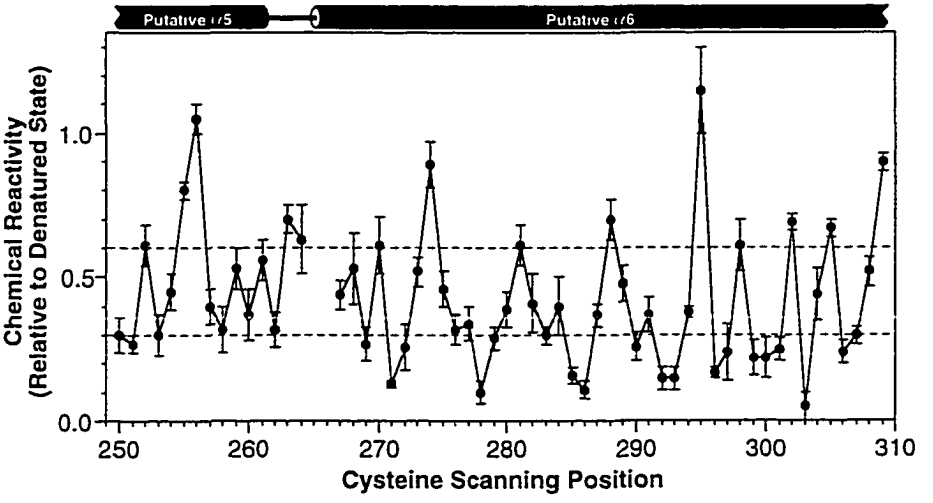


Figure 3.4. Cysteine scanning analysis of chemical reactivity in the targetted region of the cytoplasmic domain. The chemical reactivity of each engineered cysteine was determined as described in the legend of Figure 2, yielding the closed symbols. Highly exposed positions are defined as those possessing a relative chemical reactivity above 0.6 (upper dashed line), while highly buried residues are those exhibiting a chemical reactivity below 0.3 (lower dashed line). Also shown (top) are the predicted extents of the putative helices α5 and α6 within the targetted region (see Discussion). The gap in the data (positions 265, 266) indicates cysteine subsitutions that block receptor expression.

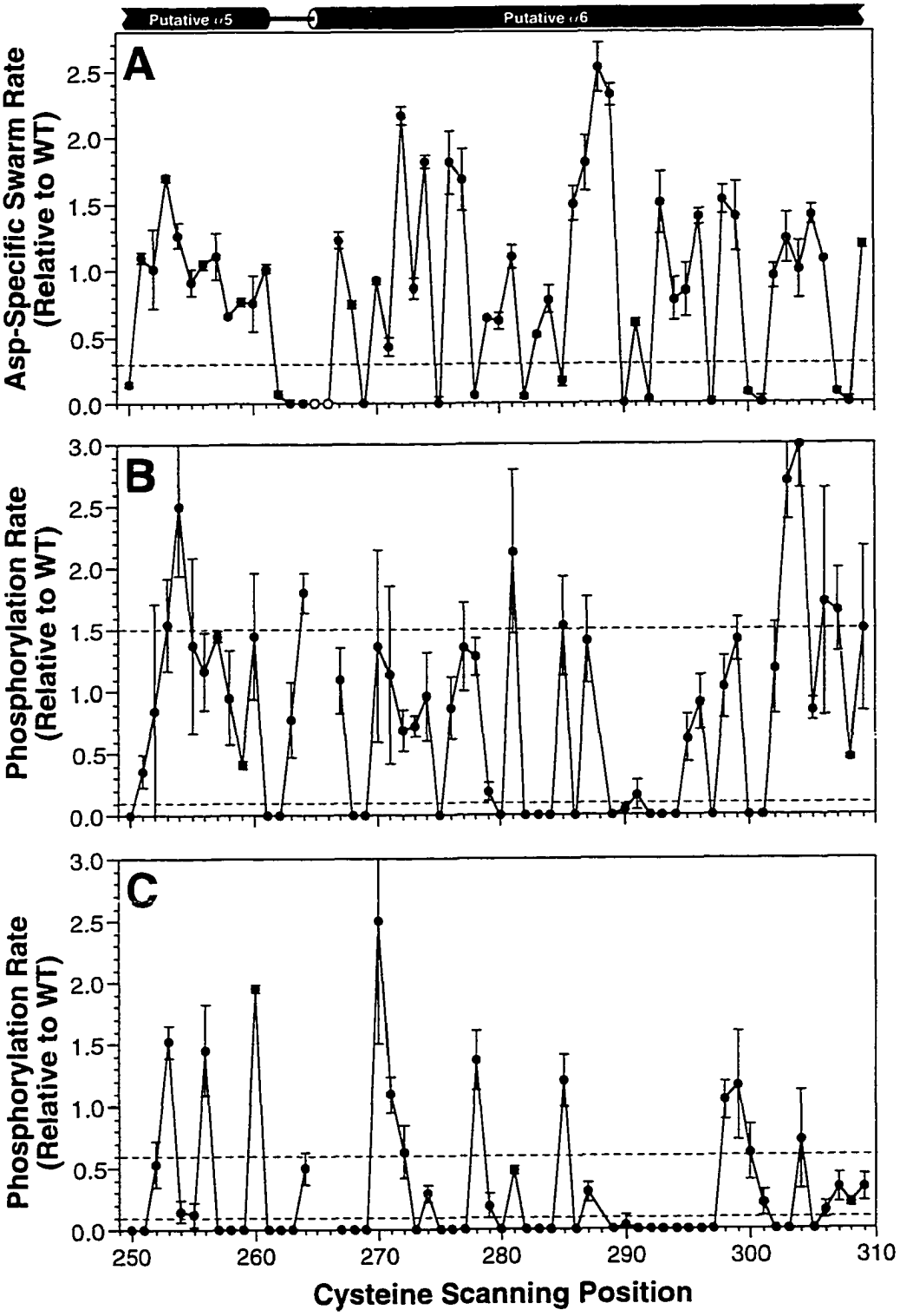
reactivities above 0.6, while buried positions exhibit reactivities below 0.3 (indicated by the upper and lower dashed lines in Figure 3, respectively). Positions 270 through 306 display a clear oscillation between highly reactive and unreactive sulfhydryls, with a periodicity different from that expected for a β-strand but matching that exhibited by a surface-exposed α-helix. In particular, local reactivity minima representing the most buried residues are located at residues 271, 278, 285, 292, 299, and 306, revealing a seven-fold periodicity characteristic of an α-helix involved in a coiled-coil or four-helix bundle. The same seven-fold periodicity is observed for local reactivity maxima representing the most exposed residues at positions 274, 281, 288, 295, 302 and 309, providing further evidence for α-helical secondary structure. Furthermore, when these highly exposed and buried positions are mapped onto a helical wheel, they fall on the opposite faces of an amphiphilic α-helix, respectively, such that the exposed positions lie on the charged face and the buried positions lie on the hydrophobic face (see model in Discussion, Figure 3.7A). At the N-terminal end of the scanned region, residues 250 through 264 exhibit few highly buried positions, indicating that any secondary structure elements which may be present in this region are largely solvent exposed and thus not as easily recognized by oscillating chemical reactivities.

Throughout the scanned region, the observed chemical reactivities are unchanged, within error, by the addition of saturating aspartate (1 mM). Such insensitivity to ligand indicates that the pattern of solvent exposure and burial within the region is not detectably altered by the ligand-induced conformational change. Overall, the results reveal the presence of an amphiphilic α-helix with distinct buried and exposed faces, extending at least from positions 270 through 309. Hereafter, this helix is termed the "first methylation helix" to distinguish it from a second methylation segment also located in the cytoplasmic domain. The exposed face of the first methylation helix includes three regulatory methylation sites at positions 295, 302, and 309, while the opposite, hydrophobic face is buried in both the apo and aspartate-occupied states of the receptor.

Identification of Side Chains Critical for Receptor Function *in vivo*. The effect of each cytoplasmic cysteine substitution on receptor function *in vivo* was measured to identify critical side chain positions in the working, fully assembled receptor-kinase complex. The plasmid pSCF6 encoding a wildtype or engineered receptor was transformed into the *E.coli* strain RP8611, which lacks chemoreceptors but contains an otherwise complete chemotaxis system (123). Overexpression of a functional receptor restores the ability of cells to migrate or "swarm" up a self-generated concentration gradient of aspartate on semi-solid agar plates (1, 41). This relatively insensitive assay is designed to detect the most extreme receptor perturbations that largely destroy receptor function, while more subtle receptor perturbations are detected by a sensitive *in vitro* assay (see below). The present analysis focuses on the functional effects of the engineered cytoplasmic cysteines, which are presumed to exist in their free sulfhydryl state due to the reducing environment of the cytoplasm.

Figure 3.5A summarizes the aspartate-specific swarm rates observed as the mutant cysteine is scanned through the targetted region of the cytoplasmic domain. Cysteine substitutions which suppress activity to less than 30% of the wildtype aspartate-specific swarm rate are classified as inhibitory (below the dashed line in Figure 3.5A). Each of these inhibitory substitutions generates a substantial receptor perturbation that cannot be overcome by receptor adaptation or overexpression. Interestingly, all five of the cysteine substitutions from positions 262 through 266 virtually eliminate receptor function (I262C, D263C, T264C) or expression (V265C, T266C). It follows that the side chains at these consecutive positions are critical to receptor assembly, function, or stability. Just C-terminal to these positions lies the first methylation helix identified by chemical reactivity measurements, spanning residues 270 through 309. Within this helix, nine inhibitory substitutions are observed to lie on the buried, hydrophobic face (see model in Discussion, Figure 3.7B) while only two inhibitory cysteines lie on the solvent-exposed

Figure 3.5. Effect of cysteine substitutions on receptor activity. (**A**) Relative rates of aspartate-specific chemotactic swarming *in vivo*. Plotted is the aspartate-specific chemotaxis swarm rate of cells overexpressing a given cysteine-containing receptor relative to the corresponding rate of cells overexpressing the wildtype receptor (see text). Open circles indicate the two cysteine substitutions that block protein expression. The dashed line is drawn at a swarm rate of 0.3, below which the substitution is referred to as inhibitory. Also shown are the relative rates of phospho-CheY production by the reconstituted receptor-CheW-CheA ternary complex, determined by R. Bass where the engineered cysteine is either (**B**) reduced, leaving free thiols, or (**C**) oxidized to form a disulfide bond. Reactant concentrations were: 3 μM reduced receptor dimer, 2 μM CheW, 0.25 μM CheA monomer, and 10 μM CheY. The buffer was 50 mM Tris, pH 7.5 with HCl, 50 mM KCl, and 5 mM $MgCl_2$, and the reaction was initiated by the addition of [γ^{32}P]ATP to a final concentration of 0.1 mM. The relative rates use the corresponding rate of the wildtype ternary complex as an activity standard (see text). The upper dashed line is drawn at a relative rate of 1.5, above which the substitution is referred to as superactivating. The lower dashed line is drawn at a relative rate of 0.1, below which the substitution is referred to as highly inhibitory.



face. Thus, the *in vivo* functional assay reveals that the buried face of the first methylation helix is more easily perturbed by cysteine substitution.

Identification of Side Chains Critical for Receptor Function *in vitro*. Further functional analysis of the engineered receptor, carried out by R. Bass, utilized the reconsitituted receptor-kinase signaling complex, which provides a sensitive *in vitro* assay for the effect of cysteine substitutions on receptor-mediated kinase regulation (25, 41, 153). The reconstitituted complex consisted of the semi-purified receptor in isolated *E. coli* membranes to which the purified soluble components were added, including the coupling protein CheW, the histidine kinase CheA and the aspartate kinase CheY. A sufficient molar excess of CheY was used to ensure that receptor-regulated histidine kinase activity was the rate-limiting step in the formation of phospho-CheY, which was quantified by its level of ^{32}P incorporation. Under these conditions, the apo wildtype receptor activates the bound histidine kinase and the emsuing formation of phospho-CheY, while aspartate binding to the periplasmic domain of the receptor inhibits phospho-CheY formation by a factor of 10^2 to 10^3-fold (25, 41, 153). Due to this large dynamic range, the *in vitro* phosphorylation assay is highly sensitive to subtle receptor perturbations.

Figure 3.5B summarizes the effects of cysteine substitutions on kinase activation by the apo-receptor in its reduced state. Plotted in this figure are receptor-stimulated phosphorylation rates, where each rate has been converted to a specific activity by normalization to the rate observed for the same concentration of wildtype receptor. Interestingly, ten of the sixty cysteine substitutions super-activate the specific phosphorylation activity at least 1.5-fold while, at the other extreme, nineteen substitutions inhibit the phosphorylation activity over 10-fold. Within the first methylation helix identified by chemical reactivities, sixteen super-activating and inhibitory substitutions lie on the hydrophobic, buried face, while only six lie on the exposed face (see model in Discussion, Figure 3.7B). Thus, the *in vitro* assay of

receptor-mediated kinase activity indicates that the buried face of the first methylation helix is more easily perturbed by cysteine substitution than its exposed face, although the observation of several perturbing substitutions on the exposed face shows that this helix surface is also important.

Comparison of the *in vivo* and *in vitro* activity data (Figure 3.5) reveals several cysteine subsitutions that inhibit receptor function over 10-fold in both assays (residues I262, R269, I275, I282, S290, R292, A297, I301). Most of these perturbations are associated with non-conservative substitutions that might disrupt receptor structure or its interaction with other proteins, such as cysteine for isoleucine, arginine or glutamate. Interestingly, all three cysteine for isoleucine substitutions block receptor function and yield intermediate solvent exposures in the chemical reactivity assay (Figure 3.4). In the native receptor, these isoleucines may be partially solvent exposed or they may reside in highly buried packing interfaces that are partially disrupted by cysteine substitution. The latter type of structural perturbation, if present, is not sufficiently common to interfere with secondary structure identification. For example, many highly buried cysteines are observed on the hydrophobic faces of the control helix α2 and the first methylation helix (Figs. 3.3-3.5, and the model in Figure 3.7 below)

Identification of Functionally Important Helix-Helix Contacts. The cysteine scanning results demonstrate the existence of the first methylation helix spanning at least positions 270 through 309, and illustrate the sensitivity of its buried surface to mutagenic perturbations. Cysteine scanning alone, however, could not identify the structure(s) against which the amphiphilic helix is packed to generate its buried surface. To ascertain whether the packing interaction might lie at the interface between subunits, a disulfide mapping study was undertaken by R. Bass, making use of pairs of symmetric cysteines found within each molecule of the engineered, homodimeric receptor. Previous studies have successfully used disulfide formation rates or extents to map out helix-helix contacts or to detect thermal collisions between distal cysteines (32, 33, 61, 62, 90, 91,

115, 131, 158). Suprisingly, mild oxidation of receptor-containing *E. coli* membranes, followed by detection of disulfide-linked dimers by a standard gel shift assay yielded at least 90% disulfide formation in each of the 58 engineered receptors. Attempts to measure initial rates of disulfide formation formation were unsuccessful because the rates were too fast to measure under standard reaction conditions (each rate constant was ~ 1 $sec^{-1}molecule^{-1}$ or faster). As a result, differing rates of disulfide formation could not be resolved to identify proximal and distal cysteine pairs. These findings indicate that the cytoplasmic domain of the membrane-bound receptor possesses an unusually dynamic structure in which even distal cysteine pairs can rapidly collide and form a disulfide bond. Such extensive backbone fluctuations are consistent with the dynamic structure previously observed for the isolated cytoplasmic domain in solution (179).

Although the unusually dynamic nature of the scanned region prevented the use of disulfide formation rates to probe structure, an inter-subunit packing face was nevertheless detected by a function-based scanning approach. The successful strategy, termed "disulfide scanning", measured the effect of each inter-subunit disulfide bond on receptor function as previously illustrated by analogous studies of the periplasmic and transmembrane helices (39, 41, 61, 91, 117). Briefly, oxidized membranes containing disulfide-linked receptors were incubated with CheA, CheW, and CheY to reconstitute the receptor-kinase signaling complex, then the ability of each receptor to regulate kinase activity *in vitro* was measured as described earlier. In effect, this disulfide scanning experiment systematically moved a unique inter-subunit disulfide bond through consecutive positions of the targetted region and determined the functional impact of a covalent constraint at each position. Of the fifty-eight disulfide bonds examined, three were found to lock the receptor in the kinase-activating state as illustrated in Figure 3.6. These receptors contained a disulfide bond at position 278, 285, or 300 and retained 25% to 140% of the wildtype kinase activation, both in the apo state and in the presence of 1 mM aspartate. Separate binding measurements confirmed that 1 mM aspartate was

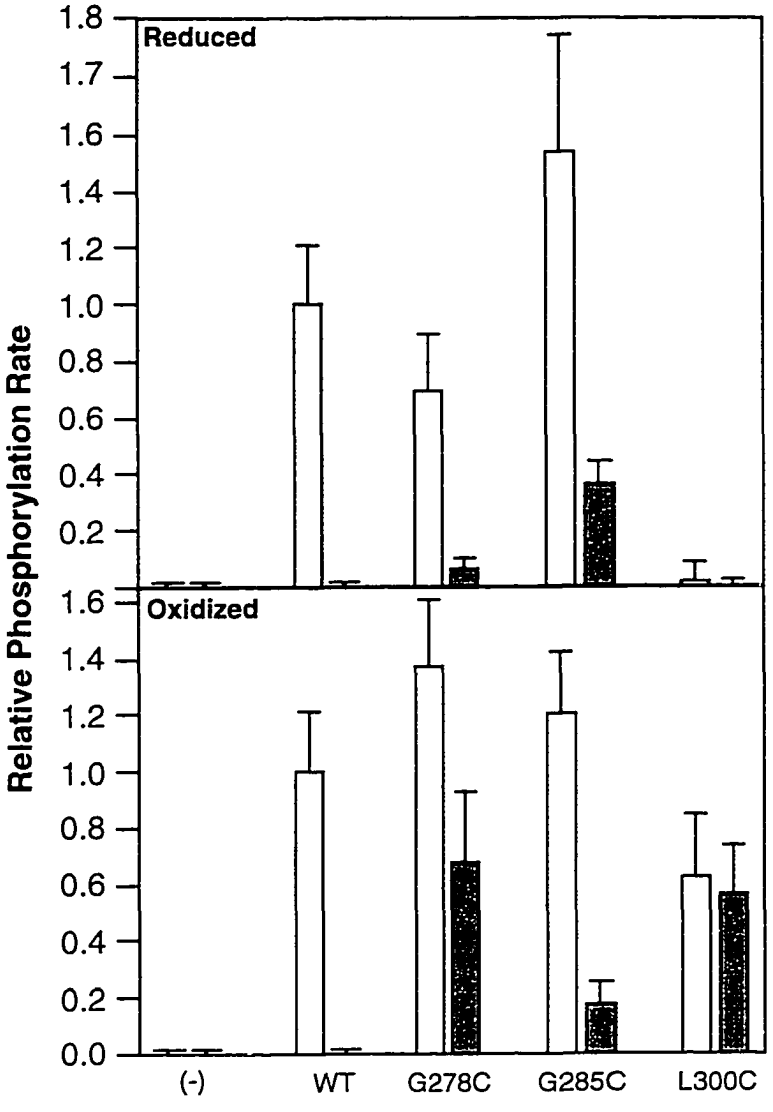


Figure 3.6. Effect of lock-on disulfides on receptor-mediated kinase regulation. *In vitro* activity was assayed as described in Figure 3.5B by R. Bass. Shown are the relative rates of phospho-CheY production in the absence (open bar) and presence (shaded bar) of 1 mM aspartate. Assays utilized the indicated oxidized receptors in which inter-subunit disulfide formation was driven to completion (see experimental procedures). No disulfide formation was observed for the wild type receptor, which lacks cysteines.

sufficient to saturate each of these receptors with bound ligand (Kim and Falke, unpublished results). Thus, the three "lock-on" disulfide bonds appear to rigidly constrain or trap the cytoplasmic domain in its kinase-activating state, regardless of ligand occupancy. Furthermore, these three lock-on disulfides all cluster to the buried,

hydrophobic face of the first methylation helix defined by chemical reactivity, thereby placing this face at an interface between two subunits in the kinase-activating state of the receptor.

More specifically, the lock-on disulfides covalently link the symmetric first methylation helices of the two subunits within the same dimer, as indicated by the following experiment. Three double-mutant receptors were generated, in which each subunit contained one cysteine at a lock-on disulfide position and a second cysteine at position 36 in the periplasmic domain (Cys36/Cys278; Cys36/Cys285; Cys36/Cys300). Previous studies of the isolated periplasmic domain have shown that upon oxidation, the interfacial Cys36-Cys36' disulfide bond forms rapidly between two subunits within the same dimer, thereby covalently stabilizing the native oligomeric structure (41, 62, 142). Thus, when a second inter-subunit disulfide is introduced, the additional crosslink will either further stabilize the existing Cys36-Cys36' covalent dimer, or will form between existing dimers to yield covalent tetramers and higher order oligomers. In practice, the standard oxidation protocol is found to convert each of the three double-mutants to a population of receptors containing over 90% covalent dimers, wherein a given dimer possesses two intra-dimer disulfide bonds (Bass and Falke, unpublished results). It follows that each lock-on disulfide forms between two subunits within the same dimer, thereby covalently trapping a native packing interaction between the symmetric pair of first methylation helices at the subunit interface.

Figure 3.5C also shows other effects of disulfide formation. The most common effect is to destroy activity. Formation of an intersubunit disulfide at 32 different positions reduced the kinase activation ability to less than 10% of wildtype activity. While it is possible that some of these disulfides lock the receptor in an aspartate-bound-like state in the absence of ligand, such an effect would be indistinguishable in this assay from disulfides which merely disrupt the native folding of the receptor. Interestingly, eight different disulfide-containing receptors not only activate the kinase to greater than 60% of

wildtype activity, but also down-regulate the kinase normally in the presence of aspartate. Five of these (T253C, S260C, G271C, S272C, and A299C) lie on the buried face of helices α5 and α6, while three (H256C, E270C, and S298C) lie on the exposed face. At the present time, the intra-dimer disulfide bonds cannot be resolved from the inter-dimer disulfide bonds because double mutants with N36C have not yet been prepared and tested.

DISCUSSION

The present study illustrates the use of cysteine and disulfide scanning to map out secondary structure and packing interactions within an unknown receptor structure, and to probe the roles of these structural features in receptor mechanism. The described cysteine scanning and chemical reactivity measurements reveal an α-helical pattern of solvent exposure for residues 270 through 309, wherein the most highly exposed and buried positions each exhibit the seven-fold repeating pattern of an α-helix involved in a coiled-coil or four-helix bundle. The observed helix is strongly amphiphilic, with the exposed and buried residues lying on opposite helix faces containing the majority of charged and hydrophobic side chains, respectively (see helix α6 in Figure 3.7A). This helix, termed the "first methylation helix", possesses three of the four receptor methylation sites on its exposed, highly anionic face. Its opposite buried face is more easily perturbed by cysteine substitutions than the exposed face (Figure 3.7B), suggesting that the packing interactions of the first methylation helix exhibit critical tolerances that modulate receptor function. Disulfide scanning identifies three lock-on disulfide bonds that place the buried helix face at the subunit interface, where it packs against a symmetric helix face provided by the other subunit in the same dimer. The exclusive association of the lock-on disulfides with this interface indicates that the observed helix-helix packing interaction is central to kinase activation. Interestingly, the presumed regulation of this interface by the transmembrane signal may involve quite subtle rearrangements since ligand binding to the

receptor does not significantly alter the pattern of solvent exposure and burial. Precedence for such subtle rearrangements exists in the periplasmic domain, which has also been found to undergo a small, aspartate-induced conformational change (40). Overall, the present results demonstrate the existence of the first methylation helix and indicate its equilibrium stability at the subunit interface of the isolated, membrane-bound receptor.

The existence of the first methylation helix explains a number of previous observations, including the apparent helical periodicity of the methylation sites, as well as the periodicity of residues involved in recognition of the methyltransferase CheR (202, 203). The importance of this region to kinase regulation had been suggested by observation that neutralization of the three regulatory glutamates, either by methylation or amidation, stimulates kinase activity by a factor of as large as 10^2- to 10^3-fold (24, 43, 195). Moreover, most of the random, second-site mutations that counteract the inhibitory effects of a charge substitution in the first transmembrane helix (A19K) are located within this region (156), where they can be presumed to modulate kinase activity. Finally, studies of the isolated cytoplasmic domain have previously implicated subunit-subunit interactions as important for kinase activation (43, 195). The present results, however, provide the first direct experimental evidence defining a specific α-helix in the cytoplasmic domain of a bacterial chemoreceptor. These results further define a specific, functionally critical helix-helix packing interaction at the subunit interface of the dimeric receptor.

Despite the equilibrium stability of the first methylation helix, the scanned region of the cytoplasmic domain is remarkably dynamic. All of the fifty-eight inter-subunit cysteine pairs examined in this region rapidly form disulfide bonds in the standard oxidation reaction, indicating that sulfhydryl-sulfhydryl collision rates are considerably more rapid than in the periplasmic and transmembrane domains (39, 41, 158). Moreover, even the most buried positions in the scanned region react at least 50-fold more rapidly with the aqueous alkylating agent than the most buried position detected in the periplasmic domain, which possesses a more static structure. Similarly, previous NMR results that

have shown the isolated periplasmic domain to be better ordered than the isolated cytoplasmic domain, which exhibits characteristic features of a molten globule (50, 179). The new results demonstrate that the noteable dynamics are not limited to the isolated domain, but are present in the full-length, membrane-bound receptor as well. It is not yet known whether the fully assembled receptor-CheW-CheA ternary complex retains these dynamics, or rather gains structural stability through quaternary contacts.

In order to determine whether the observed helix is conserved in a large class of related receptors, the aligned sequences of 56 homologous cytoplasmic domains can be compared (Chapter 4, 114). Such an alignment reveals evidence of two α-helices, termed α5 and α6, in the region corresponding to residues 250 through 309 of the aspartate receptor (Chapter 4; 114), as illustrated in Figure 3.7A. Helix α-5 lies N-terminal to the α6, while the latter helix corresponds to the region positively identified herein as the first methylation helix. Both deduced helices exhibit the heptad repeating pattern of hydrophobic and polar residues typical of helices involved in coiled-coils or four-helix bundles: **a**-b-c-**d**-e-f-g, where residues **a** and **d** are usually hydrophobic (44, 114, 129, 139, 159, 192). The **a** and **d** positions are occasionally occupied by a polar residue, particularly Ser, Thr, Asn or Gln, which can form a specific, inter-helix hydrogen bond that stabilizes a specific register of the helix-helix packing interaction (75, 76). Helices α5 and α6 are further predicted to be separated by a short bend or linker between residues L261 and V265, where a phase-shift is observed in the heptad repeating pattern of the eight enterobacterial chemotaxis receptors, including the aspartate receptor (Chapter 4; 114). Such a phase-shift is incompatible with a continuous, canonical α-helix through this region. Moreover, an alignment of more distantly related prokaryotic receptors reveals 28 examples of insertions within the putative junction, suggesting that the junction can, in such cases, accommodate a larger loop (Chapter 4; 114). Finally, flexible linker elements are often proteolytically susceptible, and it is noteable that the principal trypsin

cleavage site of the cytoplasmic domain, R259, lies in the final, presumably dynamic turn of putative helix α5 just before the α5-α6 junction begins (149).

The present findings provide strong experimental support for the existence of the putative conserved helices α5 and α6 in the aspartate receptor, as well as the junction between them. As noted above, the chemical reactivity measurements directly demonstrate that the amphiphillic first methylation helix extends from residues 270 through 309, which fall within the region identified as helix α6 by the sequence analysis (Figure 3.7A, residues 270 through 309). The evidence for putative helix α5 is less convincing, since the tested region is smaller and exhibits a weaker pattern of oscillating solvent exposure. However, all six of the highly buried or exposed residues detected in this region map to the appropriate face of putative helix α5, consistent with the existence of α-helical secondary structure. The greater accessibility of buried positions on putative helix α5 suggests that this helix is more dynamic or more solvent-exposed than α6. It should be noted that, in general, chemical reactivity measurements do not accurately define the ends of amphiphilic helices, since the helix termini can be relatively dynamic and therefore lack highly buried positions. Thus, helical regions defined by cysteine scanning may underestimate the lengths of individual helix elements.

Figure 3.7B maps out the functional perturbations due to cysteine substitutions on the putative conserved helices α5 and α6. The buried face of the first methylation helix, corresponding to α6, is especially critical for kinase regulation, since eighteen of the perturbing mutations associated with the methylation helix map to its buried face. The existence of seven perturbing mutations on its exposed face is consistent with the known importance of this face to receptor adaptation and recognition of the CheR adaptation enzyme (173, 202, 203). Interestingly, no single cysteine substitution at a methylation site destroys receptor function, an observation that lends further support to the previously proposed redundancy of the α6 methylation sites (185). The bimodal distrubution is also strong for putative helix α5, wherein all of the perturbing substitutions map to the buried

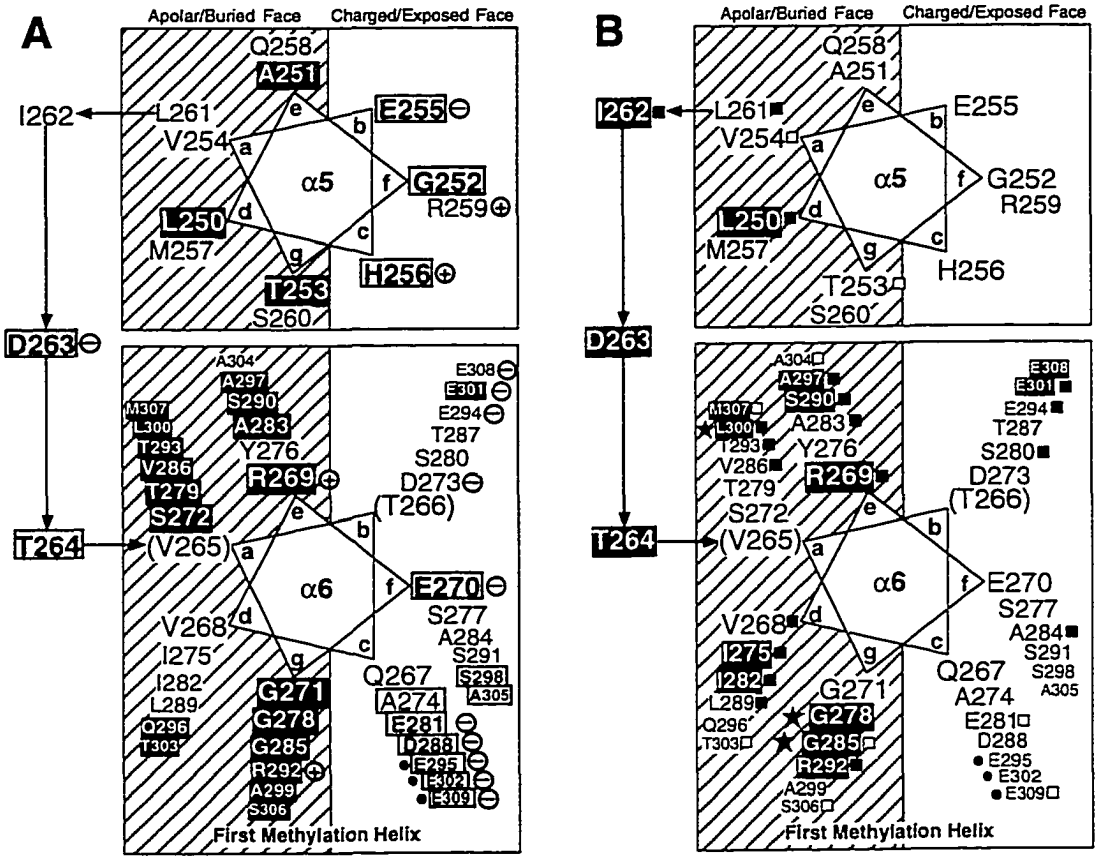


Figure 3.7. Model for cytoplasmic helices α5 and α6, displaying the experimentally determined solvent exposures and activity effects. The helices are shown with the seven-fold periodicity characteristic of coiled-coil or four-helix bundle interactions (139). (**A**) Clustering of the experimentally determined highly buried positions (black boxes) and highly solvent-exposed positions (open boxes) on opposite faces of the putative helices (exposures are defined in Figure 3.4). The hydrophobic and charged side chains are also observed to cluster on these same opposite faces, respectively. Basic (+) and acidic (-) sidechains are indicated by their charges, while the sites of regulatory methylation are highlighted by filled circles. Two of these methylation sites (positions 295 and 309) are post-translationally converted from glutamine to glutamate by CheB (97). Positions at which cysteine substitution blocks receptor expression are enclosed in parentheses. (**B**) Distribution of the experimentally determined sites of inhibitory cysteine substitution. Enclosed by black boxes are the positions where introduction of a cysteine residue inhibits chemotactic swarming *in vivo* (as defined in Figure 3.7). Smaller open squares indicate sites where cysteine substitution superactivates the kinase activity of the ternary complex *in vitro*, while closed squares indicate sites where cysteine substitution inhibits the ternary complex. Stars denote the positions of lock-on disulfides that constitutively activate the receptor bound kinase.

hydrophobic face. Most striking, however, is the distribution of perturbations in the putative linker between helices α5 and α6, corresponding to residues 261 through 265. All of the cysteine substitutions from positions 262 through 266 are found to destroy receptor signaling or expression *in vivo*, indicating that the putative junction plays an essential role in receptor function or structure. Such a junction could serve, for example, to maintain the two helices in a critical relative orientation.

Figure 3.7B also displays the locations of the lock-on disulfide bonds associated with helix α6, the first methylation helix. These inter-subunit, lock-on disulfides involve positions 278, 285 and 300, which each lie on the buried, hydrophobic face of the helix. It follows that the buried face of α6 is packed against the corresponding face of its symmetric partner α6' at the subunit interface. Additional evidence confirming this arrangement has recently been provided by independent measurements of inter-helix disulfide formation rates under weak oxidation conditions (37). Figure 3.8 presents a model for the packing interaction between two symmetric first methylation helices within the same dimer. In this model, kinase activation results from regulatory signals or lock-on disulfide bonds that stabilize the symmetric helix-helix packing arrangement at the subunit interface. The resulting packing interaction is poised for sensitive modulation, since it is stabilized by the extensive hydrophobic face of the first methylation helix but is destabilized by the highly anionic nature of its charged face. Signal-induced rearrangements of the helix-helix interface are proposed to be small, since ligand binding to the receptor has no detectable effect on the observed chemical reactivity pattern of the helix. Such findings appear to rule out major rearrangements including rotation of the α6 helix about its long axis, but are consistent with the known subtle nature of the ~1.6 Å ligand-induced conformational change in the periplasmic domain (40). Moreover, such small rearrangements could, presumably, be easily triggered by the neutralization of the anionic methylation sites. The small rearrangements could include a minor shift of helix-helix supercoiling at the subunit interface, or a change in the thermodynamic or kinetic

stability of the interface. Finally, the α6 helix is likely to have at least one additional packing interface since position **b** and **e** are partially buried (Figure 3.7A). One possibility, depicted in Figure 3.8 is that the second methylation segment participates in the formation of a four-helix bundle, as proposed by prior models based on sequence analysis (Chapter 4; 114, 192)

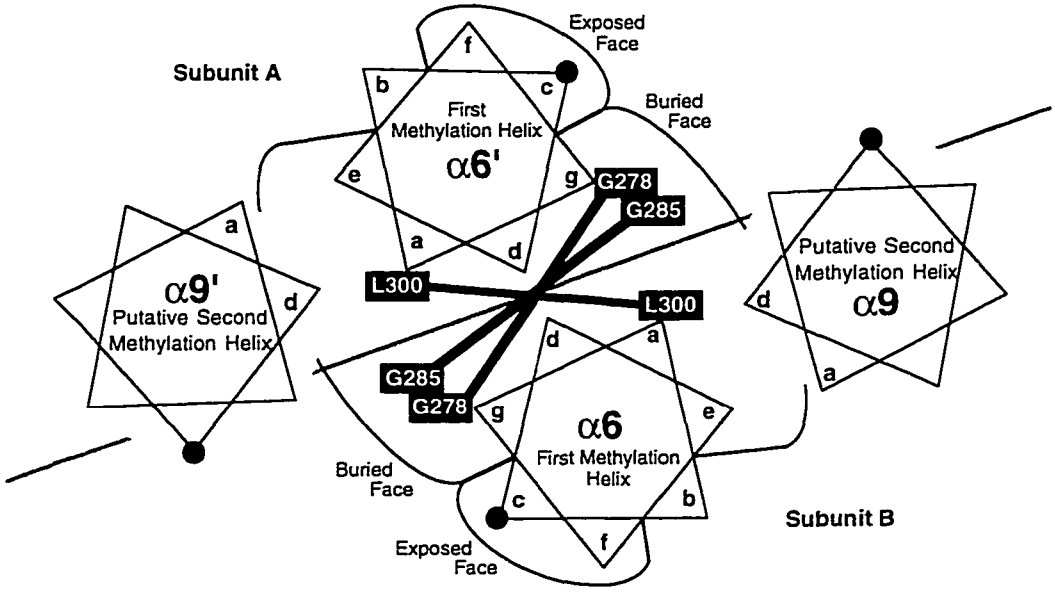


Figure 3.8. Model for the packing of the first methylation helices at the subunit interface. Shown is a view looking from the membrane toward the cytoplasm, in which the experimentally confirmed first methylation helix (α6) is oriented N-terminus to C-terminus. Positions **a** and **d** are the hydrophobic positions of the heptad repeat (see text), and the black circles indicate the positions of regulatory methylation. The inter-subunit, lock-on disulfide bonds are indicated by the symmetric helix-helix crosslinks.

Overall, the present findings indicate the usefulness of cysteine and disulfide scanning in studies of proteins not amenable to high resolution structural studies. The results provide further evidence that cysteine scanning, coupled with analysis of chemical reactivity, can map out secondary structure elements with stable exposed and buried surfaces (2, 3, 61), and that cysteine scanning coupled with activity measurements can identify functionally important residues and surfaces (67, 116, 175). When the cysteine scanning method is extended to carry out a disulfide scanning analysis, it becomes

possible to trap and detect specific contacts between pairs of residues or secondary structure elements (37, 39, 41, 61, 90, 91, 115, 117), and some of these crosslinks may trap the activated state of a signaling protein (39; also present results). Such lock-on disulfide bonds can provide significant insights into the molecular mechanisms of protein action.

Chapter Four

Modeling the Structure and the Molecular Mechanism of the Aspartate Receptor

ABSTRACT

By comparing the sequences of the various small-molecule-binding chemotaxis receptors, and with the aid of molecular modeling, insights can be obtained into ligand binding and the accompanying transmission of information to the cytoplasmic domain. Within the ligand binding site itself, negative cooperativity appears to involve rearrangement of non-coordinating sidechains upon the binding of aspartate to one site, resulting in the constriction of the other site. Non-coordinating residues also appear to provide ligand specificity by shaping the binding pocket to generate receptor-specific ligand coordination, despite the fact that different receptors use the same coordinating sidechains. The bound ligand "pushes" on the signaling helix α4/TM2, causing it to shift in a swinging piston type of motion. Such a motion is consistent with the ^{19}F NMR observed effects of aspartate binding in the 4-F-Phe-labeled ligand binding domain,as well as with lock-on and -off disulfide bonds identified in the full length receptor. Ultimately, the signaling helix transmits the signal to the cytoplasmic domain. A working model for the structure of the cytoplasmic domain can be developed by expanding on existing knowledge through secondary structure prediction. This prediction is enhanced by a sequence comparison aligning 56 of the proteins from the MTP superfamily. Conserved patterns in these proteins suggest the presence of five α-helices in each

cytoplasmic domain subunit, which in the dimer are proposed to pack into two four-helix bundles.

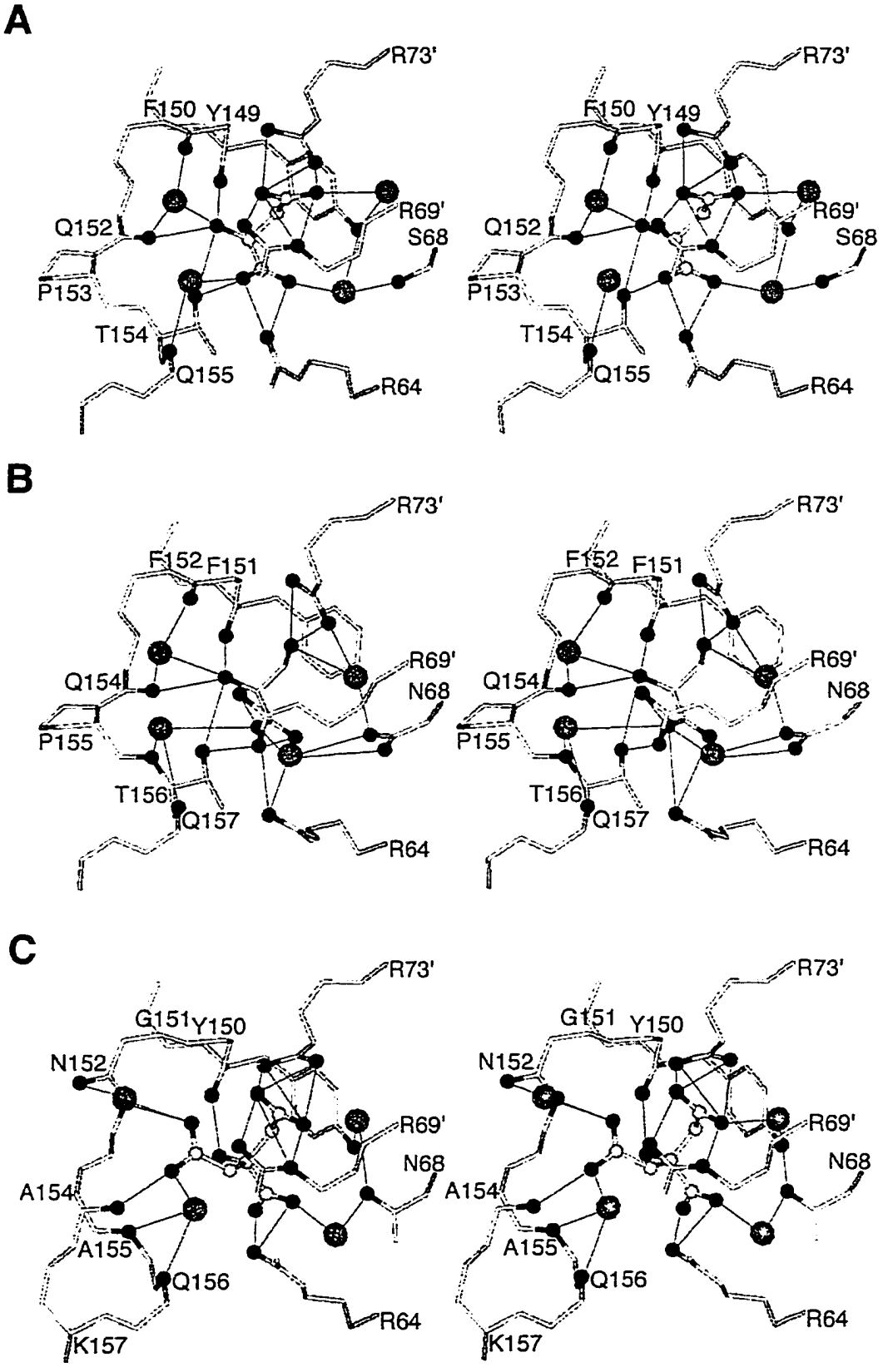
MECHANISTIC MODEL OF TRANSMEMBRANE SIGNALING

The current thesis has provided insight into the mechanism of transmembrane signaling by comparing experimental data with the known crystal structure. A careful analysis of the structure, complemented by molecular modeling raises some interesting points about ligand binding and the ensuing conformational change.

Ligand Binding Stoichiometry. The ^{19}F NMR-observed titration of the isolated ligand binding domain with aspartate (Figure 2.7) has indicated that aspartate binds to the receptor with significant negative cooperativity between the two symmetrical non-overlapping sites. This has been observed in several independent studies of the aspartate and serine receptors (22, 121, 142). Such negative cooperativity would explain how the *E. coli* aspartate receptor is able to display additive responses to aspartate and the maltose-occupied maltose binding protein (MBP) (150), since MBP binding occludes only one of the available aspartate binding sites (69). Observation of negative cooperativity in the *S. typhimurium* aspartate receptor (Chapter 2; 22, 142), as well as the serine receptor (121), neither of which interact with large-molecule ligands, suggests a further role of negative cooperativity. It is possible that asymmetry in the homodimer is important for kinase regulation. In this case it could be speculated that repellants function by stabilizing an optimal symmetric conformation.

The molecular basis for this negative cooperativity is beginning to emerge. The crystal structures of the apo and asp-occupied receptors reveal that the unoccupied site becomes constricted upon occupation of the first site (226). Moreover, ligand binding breaks a hydrogen bond between the hydroxyls of residues S68 and S68' in order to form a water-mediated contact between S68 and the bound aspartate molecule (Figure 4.1A). A study which examined the effects of point mutations at position 68 on cooperativity has

Figure 4.1. Ligand coordination by three different small-molecule receptors. Shown are stereo-views of (**A**) the structure of the ligand binding pocket of the aspartate receptor, as determined crystallographically (226), (**B**) the putative structure of the ligand binding pocket of the serine receptor, and (**C**) the putative structure of the citrate receptor. The latter two structures were modeled using the Biopolymer and Discover modules of the Insight II software package (BIOSYM), starting with the coordinates of the *S. typhimurium* receptor. In all three cases, non-coordinating sidechain atoms are represented by the sticks alone, while the atoms involved in the coordination represented by the black balls. The entire aspartate ligand is shown in ball-and-stick, with the coordinated atoms in black. Hydrogen bonds or salt-bridge interactions are represented by thin lines.



indicated that this position is indeed important to the cooperativity of aspartate binding, although the molecular basis of its role remains unclear (106). Yet, other features of the site context besides position 68 must play a role, since a mutant Asn at this position yields positive cooperativity in the aspartate receptor, while the wildtype serine receptor displays extreme (half-of-sites) negative cooperativity even though it also possess Asn at the corresponding position.

It would thus appear that multiple factors contribute to the negative cooperativity. The sequence alignment in Figure 1.4 indicates that the residues which directly coordinate the ligand are conserved, but variability is seen at the other positions in the vicinity of the binding pocket. These other positions could provide a unique context for position 68, or could themselves propagate ligand-induced conformational changes to the other binding site. In any event, it appears that no single directly- or indirectly-coordinating position alone is responsible for the negative cooperativity.

Ligand Binding Specificity. The remarkable conservation of ligand-coordinating sidechains among receptors that bind amino acids and citrate reflects similarities in the binding geometries and charge distributions of these small molecule ligands. Yet it is puzzling that different receptors apparently use an identical set of coordinating sidechains to generate high specificity for different ligands (42).

The most comprehensive way to address the specificity issue would be to compare the crystal structures of the ligand-binding domains of the various small-molecule receptors (ie. the aspartate, serine, and citrate receptors). In lieu of crystallographic data for the latter two receptors, however, one must resort to molecular modeling. Figure 4.1 shows the ligand-binding pockets of the aspartate, serine, and citrate receptors—the former from the crystallographic coordinates, while the latter two are models developed from the the aspartate receptor. These models were built by (i) replacing residues of the aspartate receptor structure with the appropriate residue from the other receptors (see the alignment in Figure 1.4) using the module Biopolymer from the Insight II package

(BIOSYM); (ii) adjusting the coordinates manually to bring about obviously favorable direct or indirect interactions, without having to alter backbone coordinates; then (iii) energy minimizing the new model to eliminate steric overlap or prohibitive electrostatic repulsions, using the module Discover (Insight II).

Comparing the structure of the aspartate receptor (Figure 4.1A) with the modeled structure of the serine receptor (Figure 4.1B), the most obvious differences in ligand coordination are in the water-mediated contacts. Because the Tyr sidechain at position Y149 is replaced with Phe in the serine receptor, the latter receptor loses a hydroxyl which coordinated a water molecule. However several new contacts are available to this water. Since R73' is now coordinating the ligand hydroxyl instead of a carboxylate, its sidechain adopts a conformation that allows as many as two contacts with the water molecule. Furthermore, the afore mentioned change of S68 to N68 enables this position to coordinate both neighboring water molecules in the serine receptor, and in addition N68' from the other subunit can now provide a hydrogen bond to this water molecule. Overall, the primary difference in coordination between the aspartate and serine ligands appears to be localized to their sidechains and associated water molecules, while their identical α-carbon amino acid carboxyl groups share very similar coordination structures. Thus minor differences in the structure surrounding the binding pocket could rearrange the conserved residues around the ligand sidechain to provide strong, multiple coordination of the hydroxyl thereby compensating for the loss of salt bridges to the aspartate carboxyl. The different site context could also contribute to serine specificity by providing a smaller binding pocket that sterically prohibits binding of the larger aspartate molecule.

The effect of the surrounding context on ligand coordination can be seen in the model of the citrate binding pocket (Figure 4.1C). Here, the α-carbon amino group of the ligand is replaced by a hydroxyl and there is an extra carboxyl, while two other carboxyls correspond to the two carboxyls of the aspartate ligand. The binding pocket is quite different at two non-coordinating positions, where (i) phenylalanine (F150 in the aspartate

receptor) is replaced by glycine (G151 in the citrate receptor), and (ii) proline (P153 in the aspartate receptor) is replaced by alanine (A154 in the citrate receptor). These changes subtly alter the backbone such that an additional carbonyl (backbone carbonyl of N152) can now form a hydrogen bond with the neighboring water molecule, thereby allowing a repositioning of the water molecule such that it can coordinate the new carboxyl. An additional contact to this carboxyl is provided by the addition of a lysine to the top of helix α4 (K157 in Figure 4.1C). This sidechain can double back to form a saltbridge with the carbonyl.

Since both the aspartate and citrate receptors contain a hydroxyl on the aromatic residue in the binding pocket (Y149 and Y150, respectively) while the serine receptor does not (F150), one might conclude that this hydroxyl is important for coordinating a carbonyl in the sidechain position of the amino acid ligands. However, looking at a model of a different receptor which binds aspartate (*Enterobacter aerogenes* aspartate receptor in Figure 4.2B), this hydroxyl is no longer present. In fact, with a phenylalanine instead of a tyrosine in the binding pocket (F144) and an asparagine instead of a serine at position 68, the binding of the water molecule in this area seems to bear more resemblance to the serine receptor than the *S. typhimurium* or *E. coli* aspartate receptors in terms of coordinating-sidechain identity. However, the *conformations* of the receptor sidechains which coordinate the ligand sidechain carboxyl bear more resemblance to the other aspartate receptors. Thus, with the noteable exception of the addition of K157 in the citrate receptor, molecular modeling suggests that the primary determinant of ligand specificity is the shape of the binding pocket, which is controlled by non-coordinating positions. Consistent with this view is the observation that ligand specificity can be altered by replacing non-coordinating residues in the aspartate receptor (I. Kawagishi, personal communication).

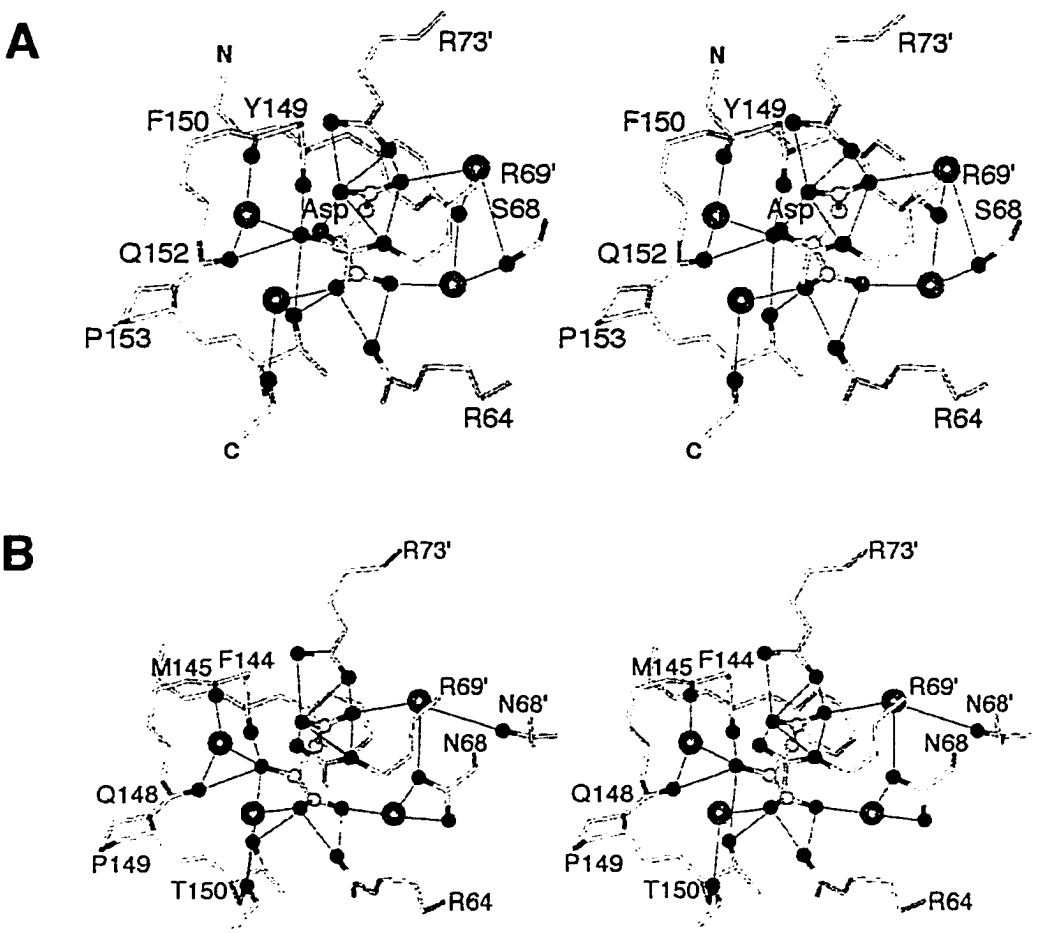


Figure 4.2. Ligand coordination by two different aspartate receptors. Shown are stereo-views of (**A**) the structure of the ligand binding pocket of the *S. typhimurium* aspartate receptor, as determined crystallographically (226), and (**B**) the putative structure of the ligand binding pocket of the *E. aerogenes* aspartate receptor, modeled using the Biopolymer and Discover modules of the Insight II software package (BIOSYM), starting with the coordinates of the *S. typhimurium* receptor. In both cases, non-coordinating sidechain atoms are represented by the sticks alone, while the atoms involved in the coordination represented by the black balls. The entire aspartate ligand is shown in ball-and-stick, with the coordinated atoms in black. The large grey balls represent bound water molecules. Hydrogen bonds or salt-bridge interactions are represented by thin lines.

The ligand induced conformational change. The ^{19}F NMR analysis of the ligand binding domain of the aspartate receptor in the current thesis has provided the first evidence that the transmembrane signal is transduced through helix α4. Subsequent studies have revealed that this is indeed the case in the intact receptor (39, 41, 90, 115, 117). A reanalysis of the crystal structure has revealed that the crystallographic data supports this model (40), in spite of the fact that the original analysis suggested otherwise (142). As shown in Figures 4.3 and 4.4, careful comparison of the the apo and aspartate-occupied aspartate structures [superimposed in the proper way (40)], reveals that the ^{19}F NMR data is highly consistent with the crystallographic data for the isolated ligand binding domain.

When aspartate binds to the apo receptor, the bulge in helix α4 must be displaced to accommodate bulk of the ligand molecule, and to properly position the coordinating sidechains of the binding pocket. As a result this forces the top of helix α4 to move downward and outward. Figure 4.3 shows that this ligand-induced conformational change changes the packing of the *para*-carbon (which is the position labeled with fluorine in the ^{19}F NMR study) of residue F150, while the *para*-carbon of position 140 remains fully solvent exposed. Consistent with this is the fact that the ^{19}F NMR resonance of F150 shows a large aspartate-induced shift, while the F140 resonance is unaffected by aspartate.

Figure 4.4 shows the long-range effects of aspartate-binding. With the *para*-carbon position of residue F107 packing directly in the interior of the four helix bundle of a single subunit (Figure 4.4A), this residue is in a position to detect the intrasubunit motion of any helix relative to the other three. Figure 4.4A also shows that the closest residues of the other subunit are quite far away from F107, so it is highly unlikely that intersubunit conformational changes, if they occur, would affect the NMR resonance of ^{19}F-labled F107. Thus the aspartate-induced change in this ^{19}F NMR resonance indicates

an intra-subunit conformational change, though it does not suggest what form that change would take. A more refined interpretation can be drawn, however, from the NMR resonances of ^{19}F-labeled F40 and F180. The former was unaffected by aspartate, while the latter shows an aspartate-induced change. Figure 4.4B shows that the *para*-carbon of position F40 packs in a static inter-subunit interface, while F180 is attached to the mobile α4. Thus it can be concluded that the crystallographically observed mobility of helix α4 and the static nature of the intersubunit face hold true in solution.

There are two ways of modeling the F30 residue. In the crystal structure of the truncated ligand-binding domain, the N-terminus of helix α1 unravels and position F30 folds back to pack in the subunit interface (Figure 2.1). However, evidence suggests that in the intact receptor, the helix continues uninterrupted throughout the membrane, which the modeled structure in Figure 4.4C shows would place the sidechain on a lipid-exposed

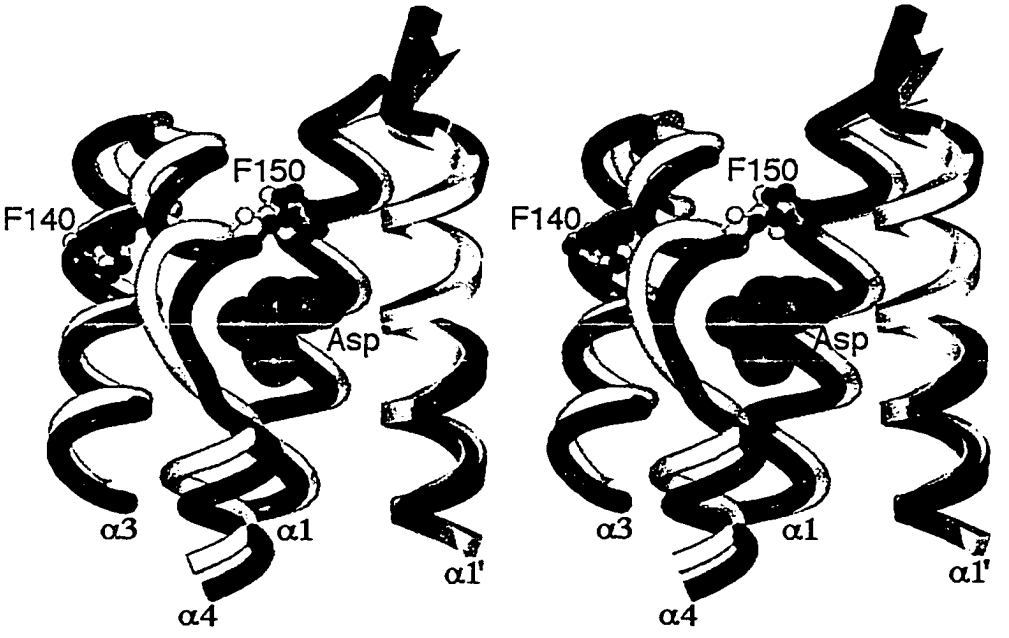
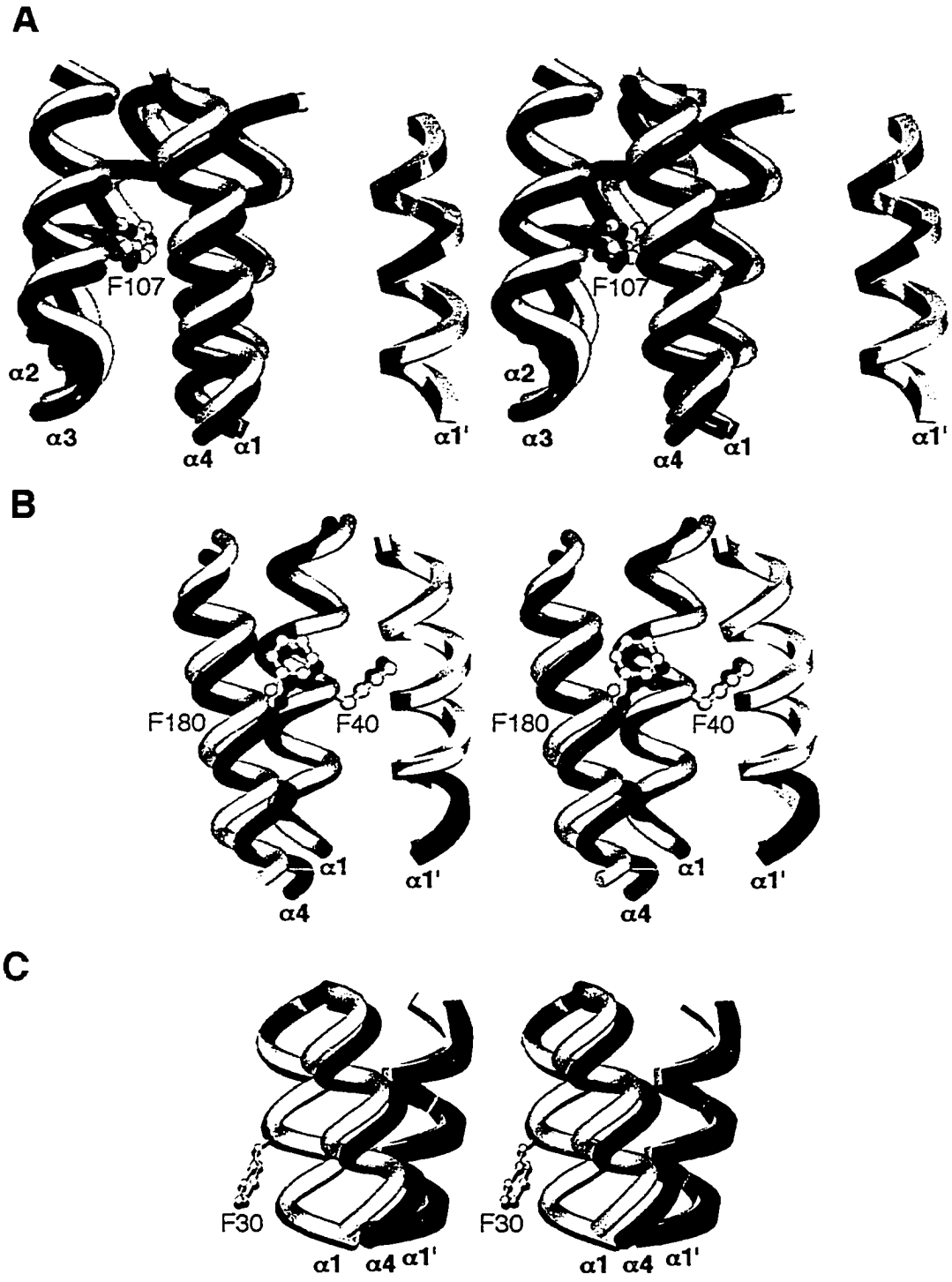


Figure 4.3. Effect of aspartate on the ligand binding pocket of the aspartate receptor. Shown is a stereo-view ribbon diagram of the aspartate binding pocket in the absense (white ribbon) and the presence (black ribbon, with the signaling helix in red) of aspartate (green CPK). The phenylalanine residues which were labeled in the ^{19}F NMR study of Chapter 2 are shown in yellow (apo) or dark blue (aspartate-bound) ball-and-stick. Subunit A is represented by round ribbons, while subunit B is represented by square ribbons.

Figure 4.4. Long range effects of aspartate on the structure of the aspartate receptor. Shown is a stereo-view ribbon diagram of three different regions of the ligand binding domain (**A**, **B**, and **C**, in order from distal to proximal to the membrane) in the absence (white ribbon) and the presence (black ribbon, with the signaling helix in red) of aspartate. The phenylalanine residues which were labeled in the ^{19}F NMR study of Chapter 2 are shown in yellow (apo) or dark blue (aspartate-bound) ball-and-stick. Subunit A is represented by round ribbons, while subunit B is represented by square ribbons.



face. Thus in the isolated ligand-binding domain in solution this residue is either packed in the dimer interface or is fully solvent exposed. In either case one would not expect to see an aspartate-induced effect on the NMR resonance of ^{19}F-labeled F30 and, in fact, none is seen.

A significant feature of this conformational change is its subtle magnitude. A model of the first transmembrane helices of the receptor (α1/TM1, α1'/TM1') based on the structure of the GCN4-p1 leucine zipper suggests that α1 and α1' would contain about 213° of superhelical twist from the C-terminal side of the transmembrane domain to the top of the periplasmic domain (178). Thus, any relative displacement of these helices would require significant re-packing and would be a high-energy rearrangement, as would any significant rearrangement of the helices within a single subunit (178). The authors thus concluded that the conformational change induced by the binding of aspartate must be a *subtle rearrangement that occurs within a single subunit.* One can thus rationalize the subtle nature of the conformational change seen in Figures 1.7, 4.3, and 4.4, since the relatively small amount of binding energy provided by the binding of a small-molecule ligand dictates that the conformational signal be small enough that the original helix-helix contacts can be maintained.

STRUCTURAL MODEL OF THE CYTOPLASMIC DOMAIN

An early step toward resolving the structure and conformational mechanism of the cytoplasmic domain of the chemotaxis receptors is the functional mapping of its sequence. Figure 4.5 shows that the sequence can be divided into discrete functional segments. *The linker region* couples the transmembrane signaling helix (α4/TM2) to the rest of the cytoplasmic domain. *The methylation region* is comprised of two distinct segments that possess the post-translational modification sites which give it its name. Finally, *the signaling domain* is the segment which directly interacts with the coupling protein CheW and the histidine kinase CheA.

Specific aspects of the structure of the cytoplasmic domain are slowly being resolved. Based on circular dichroism and hydrodynamic results, it has been known for some time that the domain is an elongated structure that is predominantly α-helical (124, 149, 220). The current thesis has provided the first direct experimental evidence showing the secondary structure of a specific segment , namely a continuous α-helix stretching at least from positions 270 to 309 as indicated in Figure 4.5. It has further defined quaternary contacts, showing that the defined helix associates with its symmetric counterpart in a parallel fashion at the homodimeric interface, at least in the kinase-activating state.

Though studies are currently under way to examine the structure of other regions of the cytoplasmic domain (Bass and Falke, unpublished; Butler and Falke, unpublished), the structure of the majority of the domain remains unresolved. Figure 4.6 presents a proposed sequence alignment for the cytoplasmic domain of 56 members of the MTP superfamily (Table 1.1). This alignment facilitates secondary structure prediction which has, in turn, yielded a more detailed working model, albeit one that remains to be experimentally confirmed. (The alignment was performed using the program CLUSTALW, with subsequent adjustments being made by hand.) The signaling domain shows far more conservation than the remainder of the domain, but pairwise sequence identities of the entire domain range from 22% to as high as 85%. Though the extent of sequence identity decreases outside of the signaling domain, there are conserved features throughout the domain. Specifically, the alignment reveals five regions displaying a repeating heptad pattern (**a**-b-c-**d**-e-f-g), in which the first and fourth residues (**a**, **d**) are typically hydrophobic, separated by regions containing insertions, deletions, or conserved Gly or Pro residues. Since such a heptad repeat pattern is typically seen in coiled-coil or 4-helix-bundle proteins (44, 129, 139, 159), these regions, highlighted on the bottom of Figure 4.5, are predicted to be α-helical. Similar predictions have been independently proposed (114, 192). Polar hydrogen-bonding side chains, sometimes found at the **a** and

Figure 4.5. Sequential map of the cytoplasmic domain of the bacterial chemotaxis receptors. Shown is a linear schematic depiction of the domain, attached to the second transmembrane (signaling) helix. Indicated within the bar representing the sequence, with the specific residue noted immediately above, are: (inverted triangle) the site of highest proteolytic sensitivity, which marks the approximate end of the linker region; (**M**) sites which have been observed to become post-translationally modified through deamidation by CheB, methylesterification by CheR, and demethylation again by CheB. The sites have been identified in the aspartate, serine, and ribose/galactose receptors (155, 169, 202). The serine and ribose/galactose receptors have an additional site of methylation, indicated by the parentheses (98, 155); (**W**) sites of second-site-revertants to disruptive mutations on CheW (122). Also, the first shaded region within the bar indicates the region which is most highly conserved among the MTP family—which is the basis upon which a protein is identified as a family member—while the C-terminal shaded region indicates the region which binds to the methyltransferase, CheR (219). Genetic studies on the serine receptor have indicated sites of lock-on (open circles) and lock-off (closed circles) mutations *in vivo* (9, 10). The open squares indicate the locations of individual second-site revertants to a deactivating mutant in the first transmembrane helix in the aspartate receptor (156). The large rectangles on the bottom indicate the regions predicted to be α-helical, based on a conserved heptad repeat pattern **a**-b-c-**d**-e-f-g, where positions **a** and **d** are typically hydrophobic (114), also Figure 4.5). The red area indicates a region in which the helicity has been experimentally confirmed (Chapter 3). The arrows below these rectangles indicate where 14-residue insertions occur within helices in a subfamily of the MTP proteins

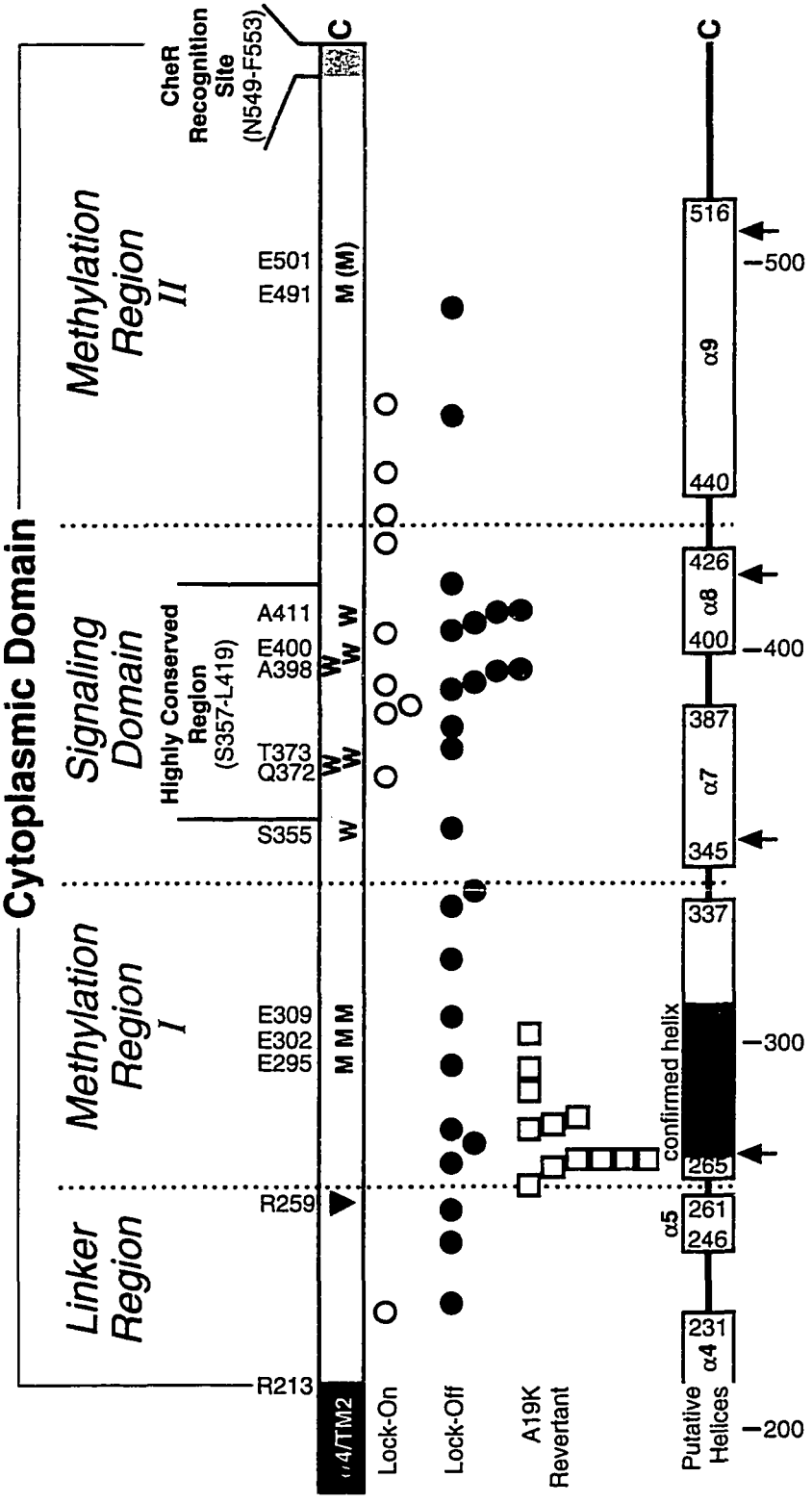
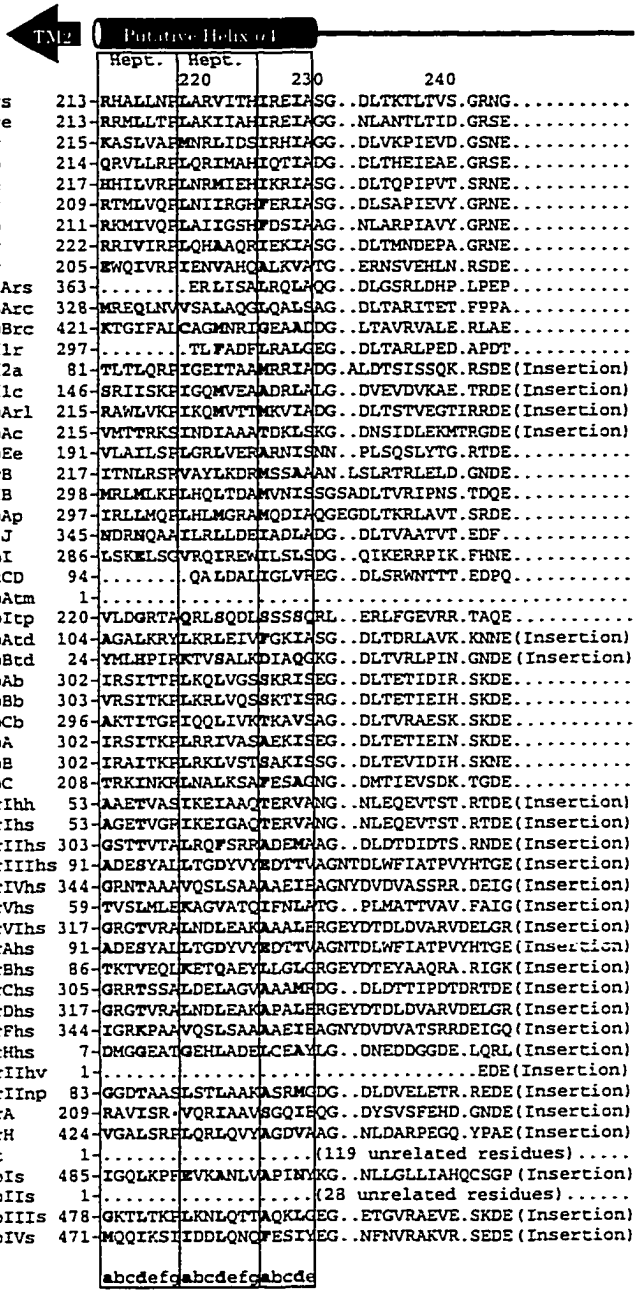


Figure 4.6. Sequence alignment of the cytoplasmic domains of the proteins of the MTP superfamily. The cylinders above the sequences indicate the regions that are predicted to be α-helical, based on the conserved pattern of seven-fold repeats: **a**-b-c-**d**-e-f-g, where **a** and **d** are generally non-polar (44, 129, 139, 159). The insert between putative helix α4 and putative helix α5 has alternatively been reported to lie between putative helix α5 and putative helix α6 (114). Either possibility seems equally plausible. The triangle above the sequences between residues 259 and 260 in the aspartate receptor indicates the site of high proteolytic sensitivity in that protein. The circles above the sequences indicate the sites where regulatory methylation occurs in the aspartate receptor (202). In the related serine (169) and ribose/galactose (155) receptors, an additional site of methylation exists, indicated by the small square above the sequences. The methylation sites generally consist of a Glx-**Glx** pair at positions b and c of the heptad repeat, where the second **Glx** residue is the one that is methylated. Positions in the putative methylation helices (α6 and α9) where the pair occurs at these at these positions with high frequency, and thus may represent a site of methylation in a particular subfamily, are enclosed by the black boxes.



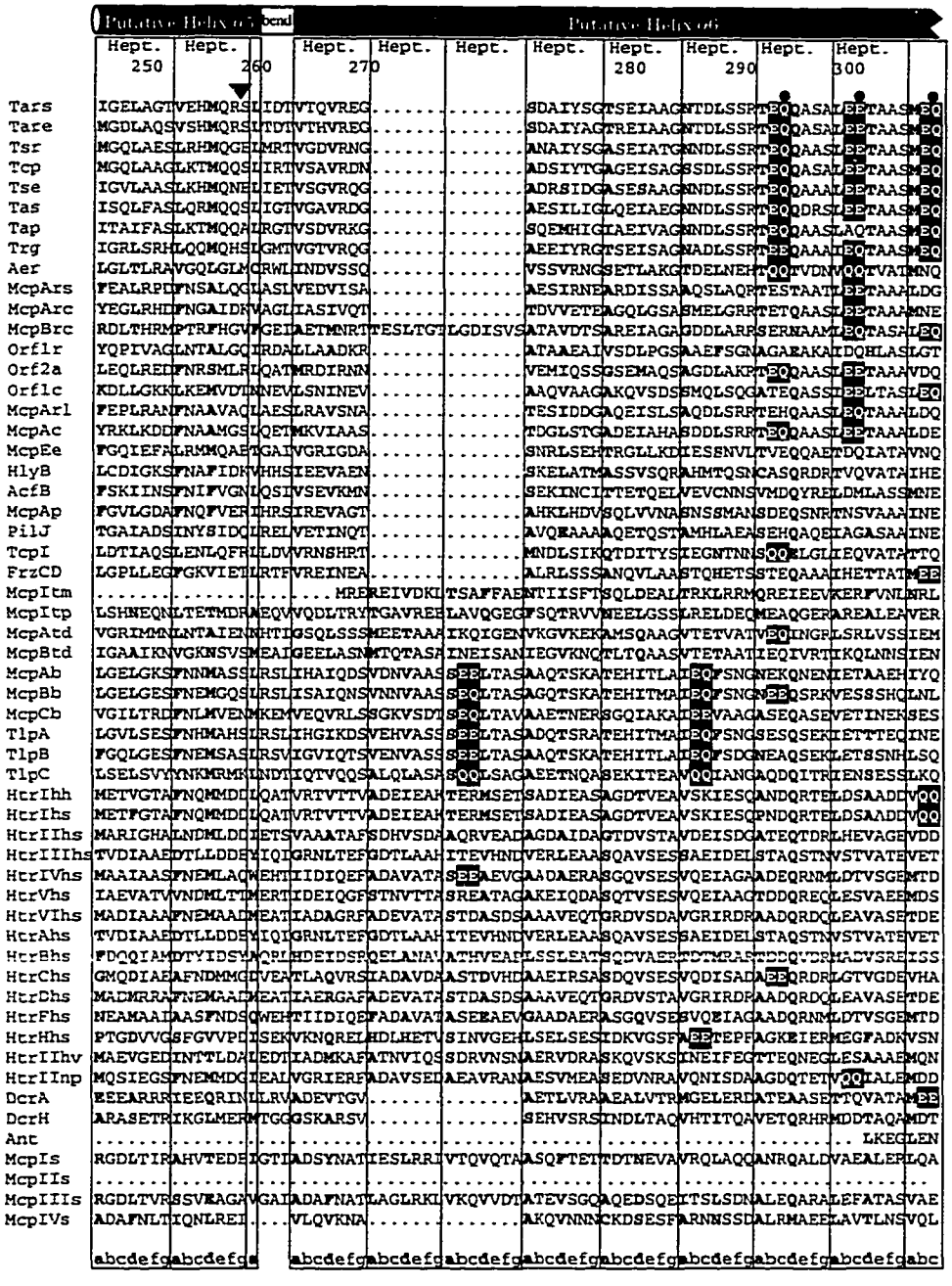


Figure 4.6. *MTP superfamily cytoplasmic domain alignment (cont.)*

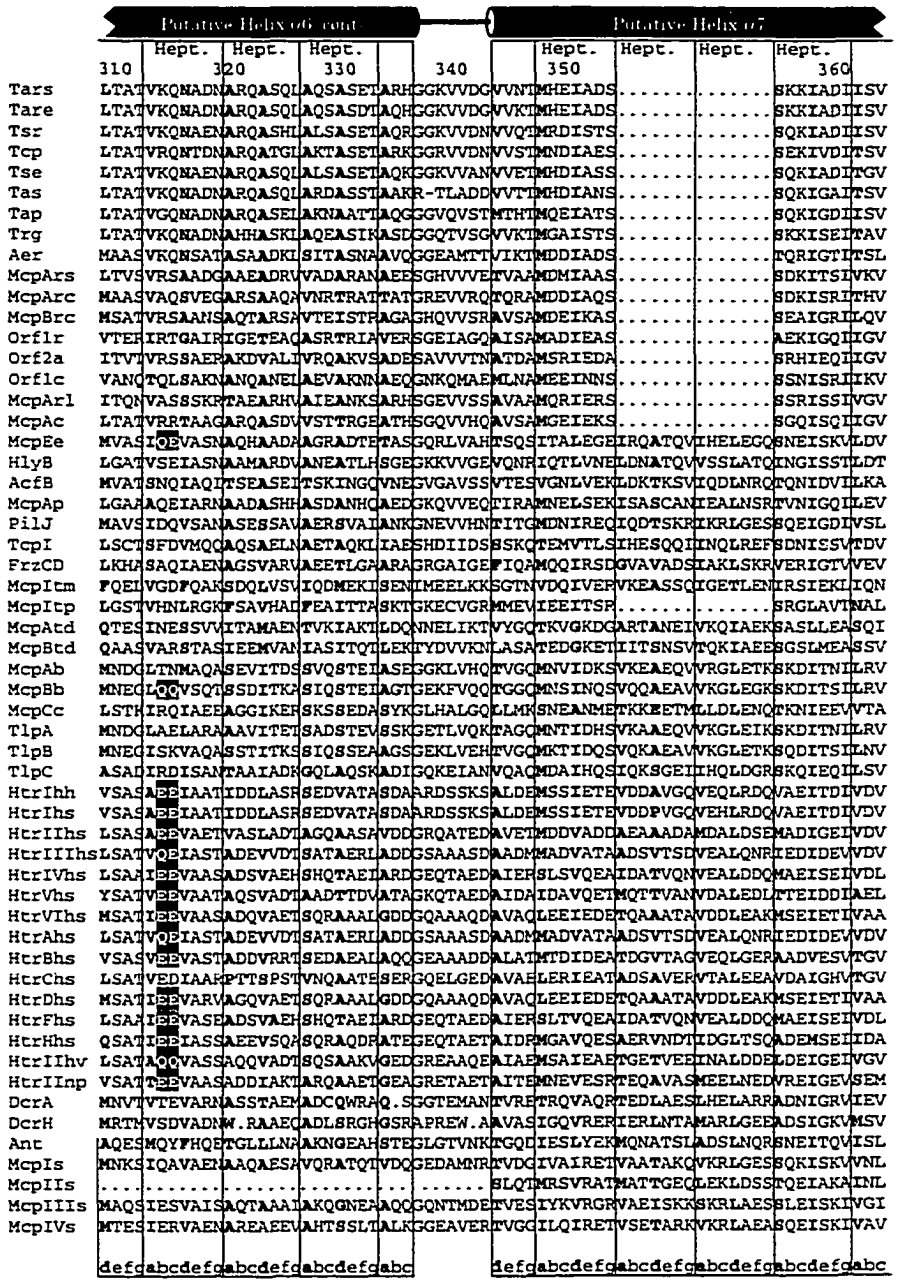


Figure 4.6. *MTP superfamily cytoplasmic domain alignment (cont.)*

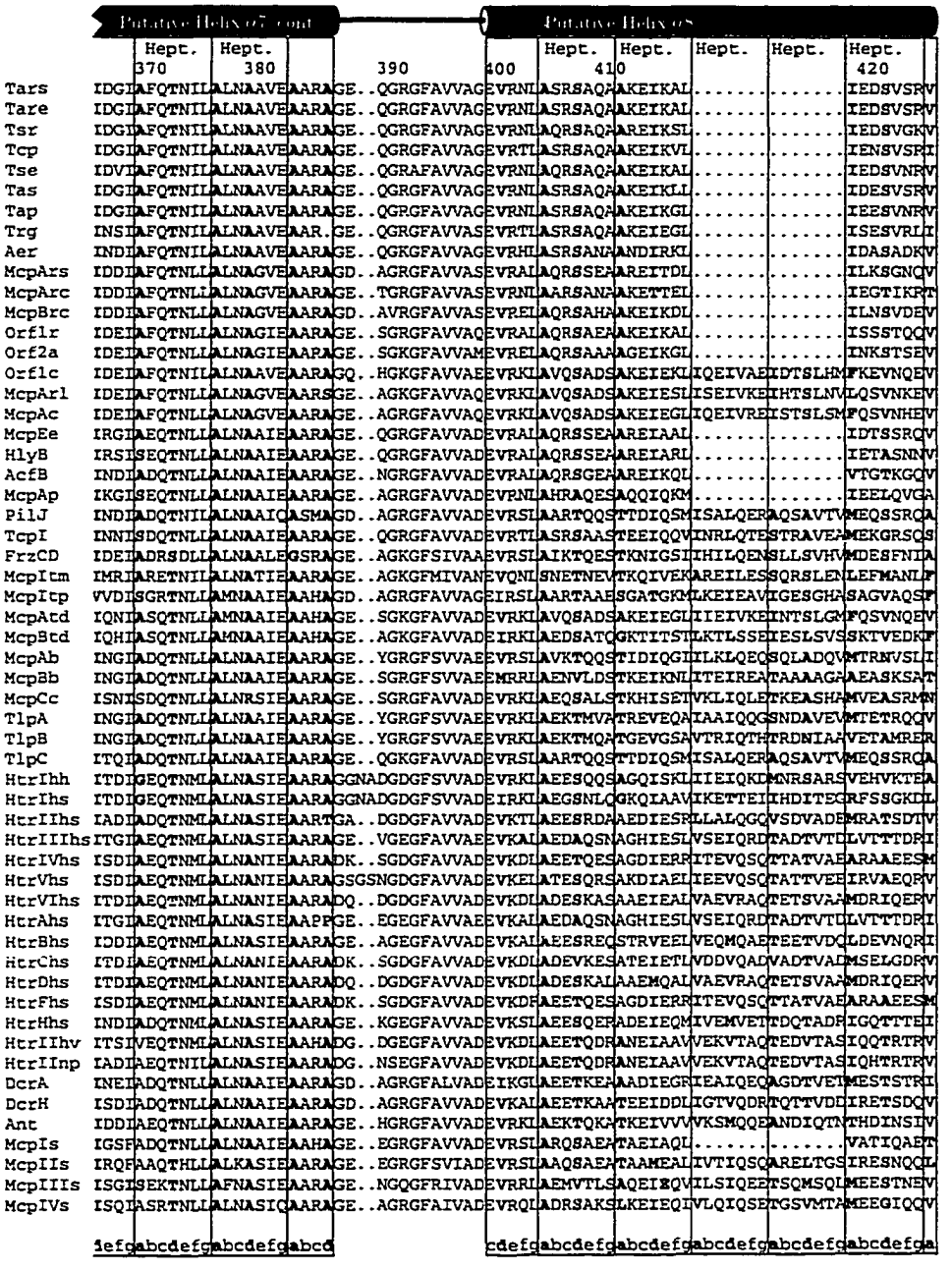


Figure 4.6. *MTP superfamily cytoplasmic domain alignment (cont.)*

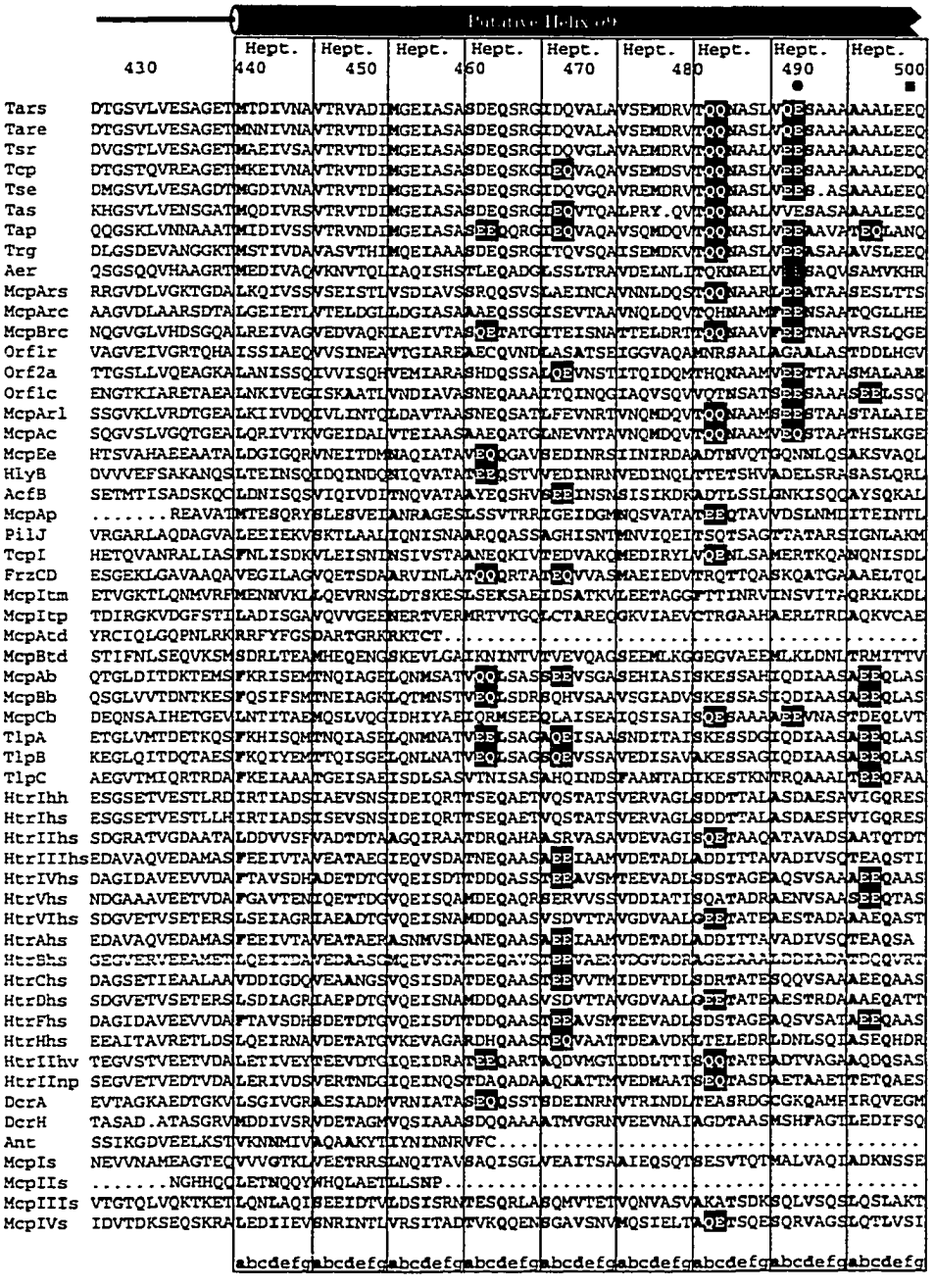


Figure 4.6. *MTP superfamily cytoplasmic domain alignment (cont.)*

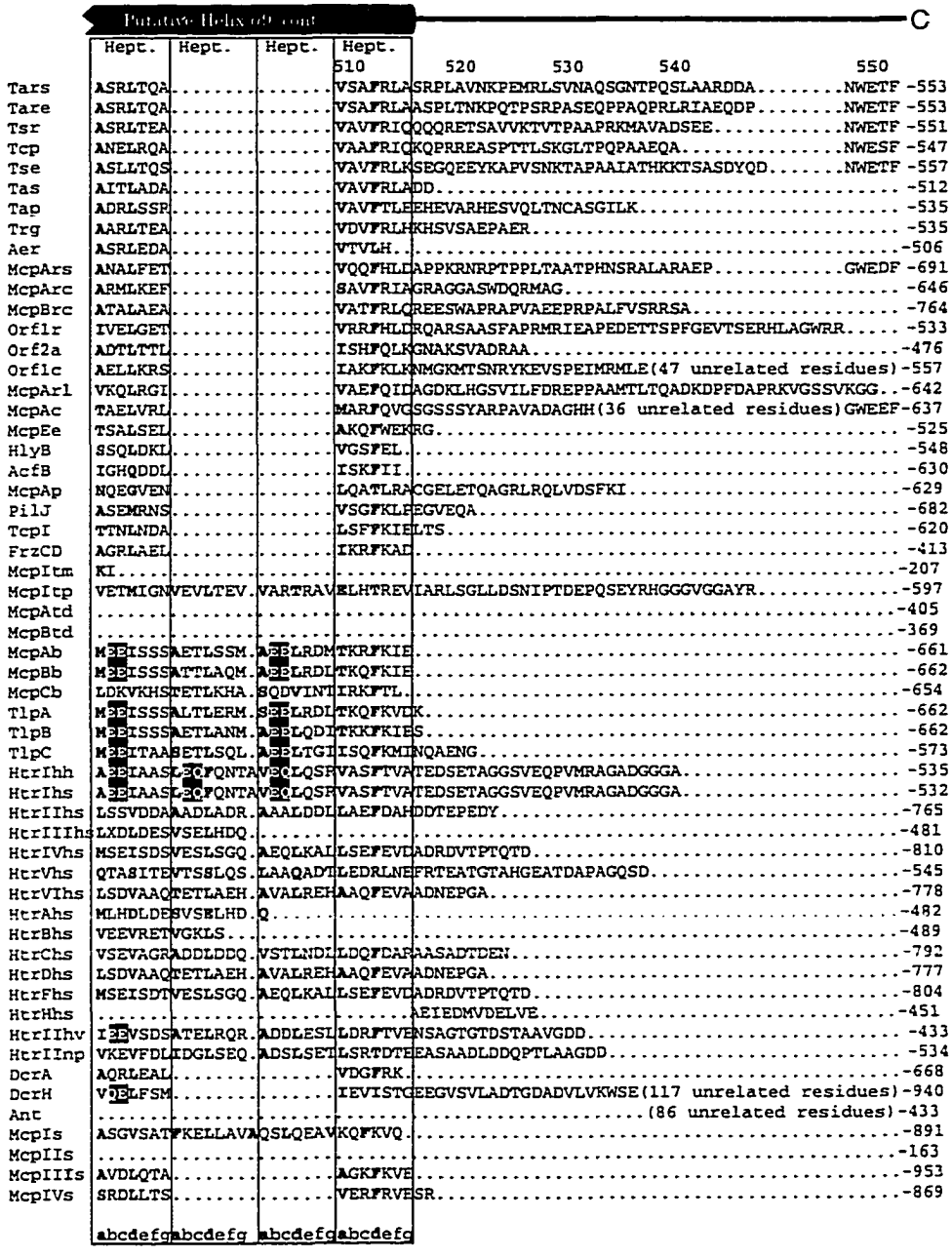


Figure 4.6. *MTP superfamily cytoplasmic domain alignment (cont.)*

d positions, help stabilize a specific register and oligomeric state of the associated helices thus minimizing structural heterogeneity (75, 76).

Figure 4.7 illustrates the results of an alternative method of structure prediction. The program NNpredict (105) employs the computational technique termed "feed-forward-neural-network." The program builds on earlier techniques which compare the input protein sequence with a database composed of the crystallographic coordinates of 105 proteins from the Brookhaven Protein Data Bank, and further improves the accuracy by searching for patterns in the sequence. For each residue in the protein, the output predicts whether the position resides in an α-helix, a β-sheet, or whether no prediction can be made. By assigning a value of 1 to residues predicted to be α-helical and 0 to all other residues, the α-helical propensity can be averaged over all 56 aligned sequences from the MTP superfamily, as shown in Fig 4.7. Although the process of averaging renders the helix propensity values arbitrary except in extreme cases, the regions which lack helix propensity stand out. Specifically, the regions with the lowest propensity, where 0 to 20% of the proteins are predicted to be α-helical, all lie within the regions predicted from pattern recognition to fall between putative helices. Thus NNpredict, as applied to the entire MTP receptor superfamily, lends support to the range of the helices suggested by the heptad repeat pattern. The current working model proposes that the ten putative helices of the cytoplasmic domain dimer fold to form the three distinct functional regions (Figure 4.8).

The Signaling Domain. The sequence alignment of the MTP receptor superfamily reveals the highest regions of conservation within the signaling domain. This conservation is not surprising, since it is this domain which is proposed to directly interact with the coupling protein and the histidine kinase (CheW and CheA, respectively, or their homologues) to regulate the cytoplasmic two-component pathway. The observed pairwise sequence identities range from approximately 40% to 100% within the signaling domain, although these identities might be somewhat inflated by the use of residues 290

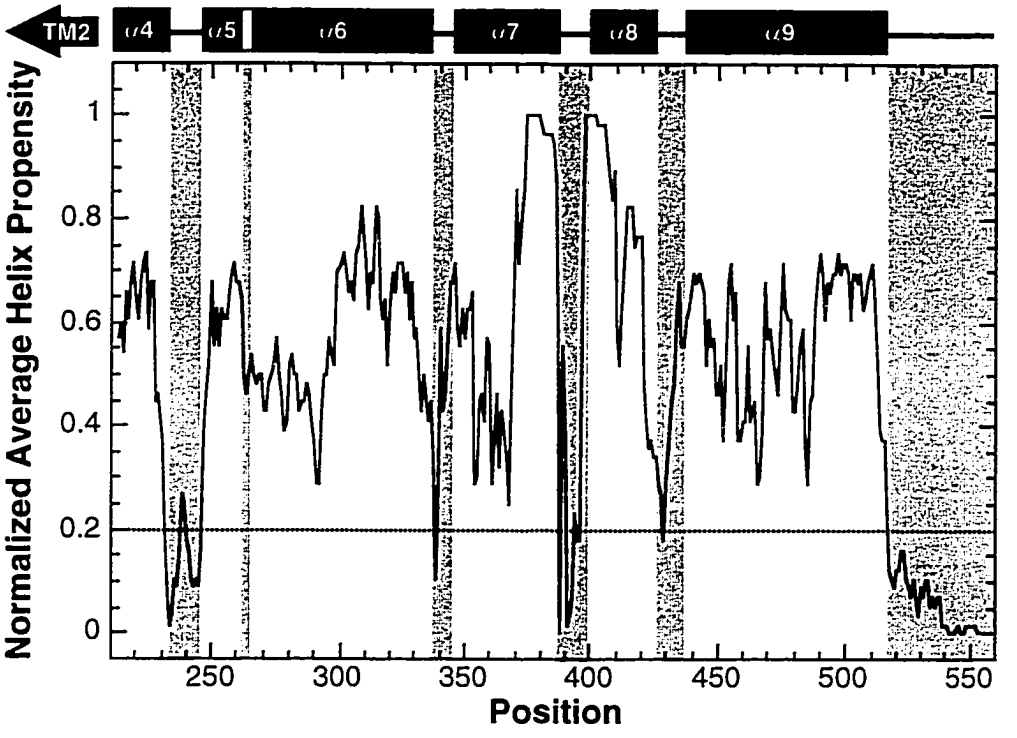


Figure 4.7. α-helical propensity of the MTP superfamily as predicted by NNpredict (105). Each protein is given a value of 1 at residues predicted to be α-helical and a value of 0 at all other positions. Thus the normalized helix propensity indicates the percent of the total number of receptors which is predicted to be helical at a given residue (numbered relative to the aspartate receptor; insertions are eliminated after determining the helix propensity, residues in gaps are given a value of 0). The boxes above the graph indicate putative helices predicted from the recognition of a heptad repeat pattern (Figure 1.5). The regions proposed to lie between these helices, indicated by the shaded boxes, display strikingly low helix propensities (ranging from 0.0 to 0.2)

to 470 for the homology search that defines a protein as a member of the MTP superfamily. Of the numerous random mutations which lock the receptor in an on or off state (9, 10), the highest density lie within the signaling domain (Figure 4.5), emphasizing the importance of this region in kinase regulation.

When the water-soluble isolated signaling domain of the serine receptor, which exhibits 87% pairwise identity to the aspartate receptor, is expressed in cells containing CheW and CheA, substantial histidine kinase regulation is seen *in vivo* and *in vitro*, indicating that the domain is an independent folding domain (11). The corresponding

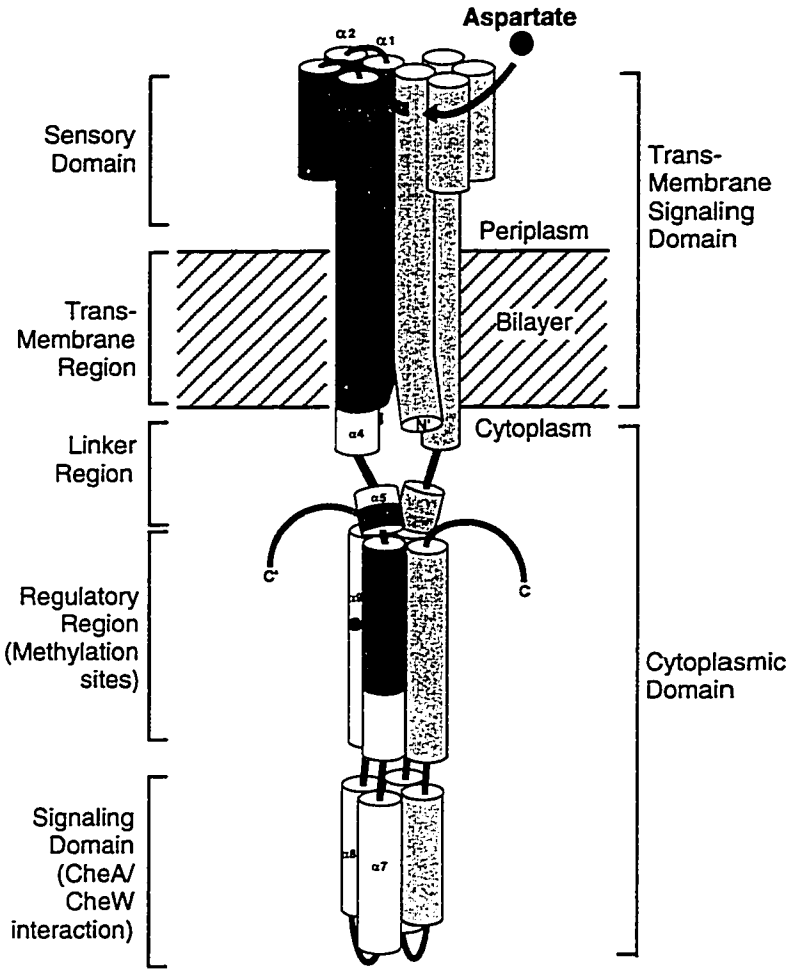


Figure 4.8. Schematic depiction of the aspartate receptor. Subunit A is shown as solid colors, while subunit B is shaded. The red cylinders represent experimentally verified α-helices (142, 158, also Chapter 3), while the blue represents an area where the experimental data, though not as strong as for the red areas, is consistent with the secondary structure prediction (Chapter 3). White and grey cylinders indicate areas predicted to form α-helices, though no experimental evidence yet exists for this prediction. The dark grey area in the sensory domain represents the aspartate binding pocket, and the sites of regulatory methylation are indicated by the black circles.

fragment in the aspartate receptor displays enhanced kinase modulation activity when it is forced to dimerize, either by inclusion of the first methylation helix, or by addition of a leucine-zipper in its place (43, 195). Thus it appears that the signaling domain contains the minimal receptor sequences required for ternary complex formation and kinase activation, and that dimerization of the module enhances kinase activation. Figure 4.5 indicates locations of second-site revertants to disruptive mutations on CheW (122), which are localized near or within the region of highest conservation. Additionally, random mutants which abolish CheW binding implicate specific regions within this region which appear to interact with the coupling protein and similar studies are being carried out to try to ascertain the specific regions of CheA interaction (J.S. Parkinson, personal communication).

Very little structural information exists for the isolated signaling domain. Circular dichroism measurements of this engineered fragment from the aspartate receptor indicate that it is predominantly α-helical (195), and similar measurements have been obtained from the signaling domain of the serine receptor (data not shown). Given this information and the high helix propensity of the regions termed putative helices α7 and α8, the simplest model for the structure of this independently folding domain is one which places the four helices α7, α7', α8, and α8' in an antiparallel four-helix bundle, as shown in Figure 4.8. Within the putative helices α7 and α8, a subgroup of proteins contains insertions, based on the pattern of appearance of non-polar residues (Figure 4.6). There are two significant features to these insertions: (i) the insertions all consist of fourteen residues, precisely the number which comprises four full turns of a coiled-coil α-helix, and (ii) with very few exceptions, each insertion in the first putative helix (α7) is offset by an insertion in the second putative helix (α8), providing support to the models which place α7 and α8 in a helical hairpin or place α7, α7', α8, and α8' in a four-helix-bundle.

The Methylation Region. The helix prediction shown at the bottom of Figure 4.5 shows that the methylation sites are proposed to lie within two α-helices. The current thesis has experimentally verified the existence of a first methylation helix. It has further revealed that this helix associates with its symmetric counterpart at the subunit interface in the kinase-activating state. Since the solvent-inaccessible surface of the first methylation helix appears larger than can be accounted for by a two-helix bundle (Figure 4.9), it would appear that another structural element packs against this helix. A model where the putative second transmembrane helix packs against the first to form a four-helix-bundle within the dimer (Figure 4.8) seems feasible for two reasons: (i) Such an arrangement would bring the four (or five) methylation sites within the same vicinity of one another, thus forming a single regulatory region in the three-dimensional structure. (ii) As is the case in the signaling domain, a subfamily of MTP proteins contains a fourteen-residue (two heptad repeats) insertion in α6 and in almost every case this insertion is offset by an equivalent insertion in putative helix α9 (Figure 4.6).

Since helix α6 associates at the dimer interface, the four-helix-bundle model dictates that putative helix α9 packs at the periphery of the α6-α6' interface. The hydropathy of the positions in the helical wheel plot (Figure 4.9) is consistent with this view. Position e of helix α6 is largely hydrophobic, albeit not so much as positions **a**, **d**, and g, thus giving helix α6 a hydrophobic face spanning four positions in the wheel. Putative helix α9, on the other hand, has only three hydrophobic positions (**a**, **d**, and g) while the other four are highly polar, suggesting a less extensive packing face, as illustrated in Figure 4.9. Also consistent with the packing model in Figure 4.9 is the observation that formation of disulfide bonds between engineered cysteines in putative helices α9 and α9' requires stronger oxidation conditions than for bonds between engineered cysteines in α6 and α6' (Bass and Falke, unpublished).

In such a model, one might expect helix α6 to be of greater functional importance than α9. Figure 4.5 illustrates several lines of evidence suggesting that this is so: (i) This

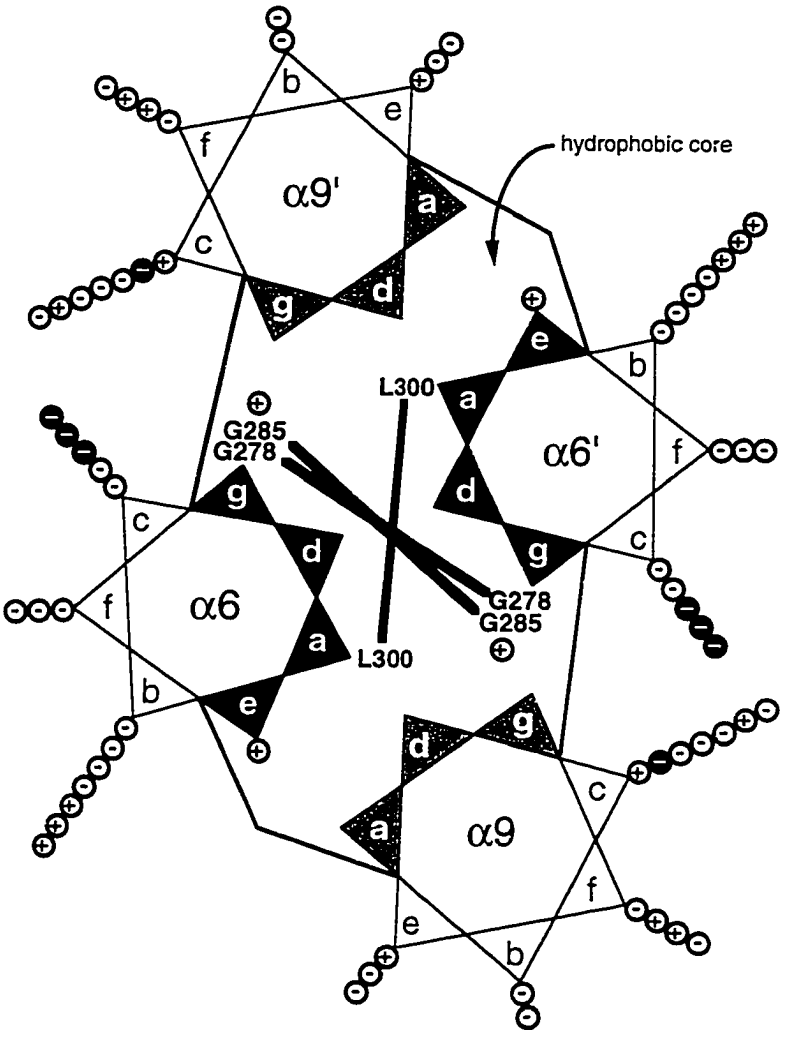


Figure 4.9. Proposed packing of the helices in the methylation region. The helical wheel plots are shown with the 7-fold periodicity of coiled-coils. Most of helix α6 has been experimentally identified, while the existence of putative helix α9 remains unconfirmed. The black spokes of helices α6 and α6' are positions on the wheel which display low solvent exposure (Chapter 3), while the grey spokes on putative helices α9 and α9' represent highly hydrophobic positions. The positions of engineered "lock-on" disulfides are indicated by the dark lines. Each occurance of a charged residue at each position on the wheel is indicated by the circled negative sign for acidic residues or positive sign for basic residues. The white negative signs in the black circles indicate the positions of regulatory methylation.

segment provides a direct conduit from the transmembrane signaling helix, by way of the linker, to the signaling domain, and thus lies on the path of information transfer from the transmembrane signaling elements to the associated kinase. (ii) This segment contains the

majority of the receptor's sites of methylation (202). (iii) Second site repressor mutations to the deactivating mutant A19K, which lies in the first transmembrane helix, for the most part cluster within this region (156). (iv) In studies of the homologous serine receptor, many point mutations which lock the receptor in a particular signaling state cluster within this region, whereas fewer such mutations have been identified in putative helix α9 (Figure 4.5) (9, 10).

Biochemical evidence also suggests that α6 is more vital to activity than α9. Truncation of putative helix α9 appears to have little effect on the ability of the isolated cytoplasmic domain to activate the kinase (195), though it should be noted that the necessity of this segment to the function of the intact receptor is unknown. In contrast, removal of the putative helix α6 destroys the kinase activating ability of the cytoplasmic domain, though replacement of this segment with a leucine-zipper sequence can restore the activity (43, 195).

Though structure prediction cannot provide a reliable indication of the vertical register of putative helix α9 compared to helix α6 ("vertical" in this instance refers to the long axis of the helices), two observations combine to suggest the preliminary model where the N-terminus of helix α6 aligns with the C-terminus of putative helix α9. (i) Helix α6 (in its putative entirety) and putative helix α9 are approximately the same length. (ii) Sequence alignment algorithms as well as visual analysis suggests that the fourteen residue subfamily insertion in helix α6 occurs seven residues from its N-terminus, while the corresponding insertion in putative helix α9 occurs seven residues from its C-terminus. Thus, in the simplest model, the N- and C-termini of the helices α6 and α9 are aligned, placing the sites of fourteen residue insertions at adjacent locations to yield an antiparallel 4-helix bundle. Needless to say, this arrangement, while providing a good starting point for future modeling, is entirely speculative and calls for experimental verification.

The sites of methylation occur on specific glutamate residues, which are either directly encoded in the gene or are produced by the CheB-catalyzed deamidation of glutamine. Where they have been identified, the methylation sites generally consist of a Glx-**Glx** pair at positions b and c of the heptad repeat, where the second **Glx** residue is the one that is methylated. Fig 4.6 shows that the location of the conserved consensus sequences of the methylation sites appears to be specific to three families. Though the specific sites of methylation have only been identified in the enterobacterial chemotaxis receptors (155, 169, 202), methylation has been shown to occur in the other two families, as well (81, 229). The first family includes the enterobacterial chemotaxis receptors (including the aspartate receptor) and displays 3-4 consensus sites, and one (based on the identification of the methylation sites of the serine and the ribose/galactose receptors) non-consensus site (155, 169). The second family is comprised of the chemoreceptors of the eubacterium *Bacillus subtilis*. These proteins display 3-5 conserved sites distinct from the enterobacterial sites. The photo-transducers of the halobacteria, which contain 1-3 conserved sites, comprise the third family. Interestingly, while the aspartate receptor contains four sites of methylation (202), the serine and ribose/galactose receptors contain five (155, 169), though the sequences of the three receptors in the vicinity of the fifth site are nearly identical. This suggests that the specificity of the methyltransferase CheR is directed by subtle differences in the methylation site context. The methyltransferase CheR is tightly bound in the vicinity of the methylation sites via its docking to the heptapeptide NWE(S/T)F at the C-terminus of the receptor (Figure 4.5) (219). It has also been shown that CheR docks to one dimer and acts in *trans* on another (119), indicating that the methylation sites of other receptors all reside in the same plane parallel to the membrane, where the CheR protein also resides.

The Linker Region. The linker region is proposed to extend from the putative end of helix α4 to the beginning of the methylation helix α6, and thus extends from residues S232 to T264 in the aspartate receptor. Perhaps the most striking aspect of the

linker region is the site of high proteolytic sensitivity which approximately marks its C-terminus (149). This sensitivity was origionally interpreted as indicating that the linker region has little or no permanent structure. Such an interpretation is counterintuitive, however. Since the short cytoplasmic extension of the N-terminal transmembrane helix (α1/TM1) has been shown to be dispensable (38), it stands to reason that the transmembrane signal propagated by the signaling helix (α4/TM2) is transmitted entirely through the covalently attached linker region to the functionally vital first methylation helix and utlimately to the associated histidine kinase. Given the rapid response of the receptor/CheW/CheA ternary complex to ligand binding (180) it is clear that there must be a pre-assembled structure which allows the propagation of a ligand-induced conformational change through the linker region to the kinase.

The sequence alignment of this region is consistent with the existence of important packing interactions within the linker region. Figure 4.6 shows conservation within this region that extends throughout most of the MTP superfamily, particularly in the region immediately following the putative terminus of α4 (Figure 4.6, this region shows low α-helix or β-sheet propensity). Furthermore, both lock-on and lock-off mutations have been identified in this region in the serine receptor *in vivo* (9, 10). Finally, receptor heterodimers in which one cytoplasmic domain has been truncated retain signaling function only when both subunits, including the truncated subunit, possess the linker region (70, 201).

MECHANISTIC MODEL FOR KINASE REGULATION

Studies of the periplasmic and transmembrane regions have resulted in a model for transmembrane signaling where the ligand induces a piston or swinging-piston displacement of signaling helix α4/TM2 (40), which spans the membrane (Figure 4.8). This displacement could either be propagated directly through the cytoplasmic domain, or could be translated into another type of displacement.

A model has been proposed involving a significant rotation of helix α6 relative to α6' (43). The solvent exposure data of Chapter 3 are inconsistent with this model because such a rearrangement of the packing face would result in a change in the pattern of solvent exposure and burial within the scanned positions. Thus, if the piston-type displacement is translated into another type of displacement within the linker region, the new type of motion is not a large rotational motion.

The lack of any known "lock-off" disulfides (which mimic the aspartate-bound state, rather than merely destroying activity) limits the development of specific mechanistic models from the data in Chapter 3, since we only have an indication of the structure in one state. It appears, however, that whatever type of displacement occurs is subtle. (i) Large displacements would likely alter the pattern of solvent exposure and burial of the helix, yet no change in the solvent exposure pattern was seen for helix α6 upon addition of aspartate. (ii) Several engineered disulfide-containing receptors retained the wildtype ability to activate the kinase and down-regulate that activity upon addition of aspartate. Though full interpretation of these results will require establishing whether these disulfides are intra- or inter-dimer, if one assumes that the "retain" disulfides on the buried face are intra-dimer, then the aspartate induced signal must fall within the range of motion allowed by the disulfide bond (approximately 2 Å). (iii) As discussed in *Mechanistic Model of Transmembrane Signaling* above, large rearrangements of α-helices in a four helix bundle would likely involve a re-packing which could prove energetically costly, and the binding of a small-molecule ligand may not be able to provide enough energy.

The data in the current thesis is consistent with a model involving the continuation of the piston-type of motion throughout the methylation region. Disulfides which lock the receptor in the kinase-activating state occur at the interface between α6 and α6'. It is possible that these disulfides inhibit the aspartate-induced signal by preventing such a translational movement of the helices relative to eachother. It must be cautioned however that the data are also consistent with other models. For instance, if aspartate shifts an

equilibrium between tightly- and loosely-associated subunits, then the lock-on disulfides would prevent this by covalently maintaining tight association. This would also provide a possible explanation for the effect of methylation on activity. Methylating the four glutamate positions on helix α6 and putative helix α9 (Figure 4.9) would reduce the anionic nature of these helices and therefore reduce electrostatic repulsion, thus allowing the helices to associate more tightly. Other conformational mechanisms, including a minor shift of helix-helix supercoiling at the subunit interface, are equally plausible. Thus, while the current thesis provides a good starting point for understanding the conformational mechanism of signal transduction through the methylation region, clearly further studies must be carried out to fully characterize it.

BIBLIOGRAPHY

1. Adler, J. (1966) Chemotaxis in bacteria. *Science* **153**, 708-716.

2. Akabas, M.H., Kaufmann, C., Archdeacon, P. and Karlin, A. (1994) Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the alpha subunit. *Neuron.* **13**, 919-927.

3. Akabas, M.H., Stauffer, D.A., Xu, M. and Karlin, A. (1992) Acetylcholine receptor channel structure probed in cysteine- substitution mutants. *Science* **258**, 307-310.

4. Alex, L.A. and Simon, M.I. (1994) Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* **10**, 133-138.

5. Alley, M.R., Maddock, J.R. and Shapiro, L. (1992) Polar localization of a bacterial chemoreceptor. *Genes Dev.* **6**, 825-836.

6. Allmeier, H., Cresnar, B., Greck, M. and Schmitt, R. (1992) Complete nucleotide sequence of Tn1721: gene organization and a novel gene product with features of a chemotaxis protein. *Gene* **111**, 11-20.

7. Alm, R.A. and Manning, P.A. (1990) Characterization of the hlyB gene and its role in the production of the El Tor haemolysin of *Vibrio cholerae* O1. *Mol.Microbiol.* **4**, 413-425.

8. Altschul, S.F., Gish, W., Miller, W., Meyers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J.Mol.Biol.* **215**, 403-410.

9. Ames, P., Chen, J., Wolff, A. and Parkinson, J.S. (1988) Structure-function studies of bacterial chemosensors. *Cold Spring Harbor Symp.Quant.Biol.* **53**, 59-66.

10. Ames, P. and Parkinson, J.S. (1988) Transmembrane signaling by bacterial chemoreceptors: *E. coli* transducers with locked signal output. *Cell* **55**, 817-826.

11. Ames, P. and Parkinson, J.S. (1994) Constitutively signaling fragments of Tsr, the *Escherichia coli* serine chemoreceptor. *J.Bacteriol.* **176**, 6340-6348.

12. Ames, P., Yu, Y.A. and Parkinson, J.S. (1996) Methylation segments are not required for chemotactic signalling by cytoplasmic fragments of Tsr, the methyl-accepting serine chemoreceptor of *Escherichia coli. Mol.Microbiol.* **19**, 737-746.

13. Arakawa, S. and Kuramitsu, H.K. (1997) . *Direct Submission*

14. Arseniev, A.S., Kuryatov, A.B., Tsetlin, V.i., Bystrov, V.F., Ivanov, V.T. and Orchinnikov, Y.A. (1987) ^{19}F NMR study of 5-Fluorotryptophan-labeled bacteriorhodopsin. *FEBS Lett.* **213**, 283-288.

15. Ausubel, F.M., Brent, R., Kingston, R.E., et al. (1992) *Current Protocols in Molecular Biology, vol. 1*. Greene Publishing Association, New York. pp. 1.1.2-1.1.0.

16. Barak, R. and Eisenbach, M. (1992) Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor [published erratum appears in *Biochemistry* 1992 May 19;31(19):4736]. *Biochemistry* **31(6)**, 1821-1826.

17. Baumgartner, J.W. and Hazelbauer, G.L. (1996) Mutational analysis of a transmembrane segment in a bacterial chemoreceptor. *J.Bacteriol.* **178**, 4651-4660.

18. Baumgartner, J.W., Kim, C., Brissette, R.E., Inouye, M., Park, C. and Hazelbauer, G.L. (1994) Transmembrane signaling by a hybrid protein - communication from the domain of chemoreceptor Trg that recognizes sugar binding proteins to the kinase/phosphatase domain of osmosensor EnvZ. *J.Bacteriol.* **176**, 1157-1163.

19. Berg, H.C. (1993) *Random Walks in Biology*. Ed. 2. Princeton Univ. Press, Princeton.

20. Bibikov, S.I., Biran, R., Rudd, K.E. and Parkinson, J.S. (1997) A signal transducer for aerotaxis in *Escherichia coli*. *J.Bact.* **179**, 4075-4079.

21. Biemann, H.-P., Harmer, S.L. and Koshland, D.E.,Jr. (1996) An aspartate/insulin receptor chimera mitogenically activates fibroblasts. *J.Biol.Chem.* **271**, 27927-27930.

22. Biemann, H.-P. and Koshland, D.E.,Jr. (1994) Aspartate receptors of *Escherichia coli* and *Salmonella typhimurium* bind with negative and half-of-the-sites cooperativity. *Biochemistry* **33**, 629-634.

23. Blair, D.F. (1995) How bacteria sense and swim. *Annu.Rev.Microbiol.* **49**, 489-522.

24. Borkovich, K.A., Alex, L.A. and Simon, M.I. (1992) Attenuation of sensory signaling by covalent modification. *Proc.Natl.Acad.Sci.USA* **89**, 6756-6760.

25. Borkovich, K.A., Kaplan, N., Hess, J.F. and Simon, M.I. (1989) Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proc.Natl.Acad.Sci.USA* **86**, 1208-1212.

26. Bourret, R.B., Borkovich, K.A. and Simon, M.I. (1991) Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu.Rev.Biochem.* **60**, 401-441.

27. Bourret, R.B., Drake, S.K., Chervitz, S.A., Simon, M.I. and Falke, J.J. (1993) Activation of the phosphosignaling protein CheY. Analysis of activated mutants by ^{19}F NMR and protein engineering. *J.Biol.Chem.* **268**, 13089-13096.

28. Bowie, J.U., Pakula, A.A. and Simon, M.I. (1995) The 3-dimensional structure of the aspartate receptor from *Escherichia coli*. *Acta Cryst.* **51**, 145-154.

29. Boyd, A., Kendall, K. and Simon, M.I. (1983) Structure of the serine chemoreceptor in *Escherichia coli*. *Nature* **301**, 623-626.

30. Brito, B.A., Palacios, J.M., Ruiz-Argueso, T. and Imperial, J. (1996) Identification of a gene for a chemoreceptor of the methyl-accepting type in the symbiotic plasmid of *Rhizobium leguminosarum bv. viciae* UPM791. *Biochim.Biophys.Acta* **1308**, 7-11.

31. Brooun, A., Zhang, W. and Alam, M. (1997) Primary structure and functional analysis of the soluble transducer protein HtrXI in the archaeon *Halobacterium salinarium*. *J.Bacteriol.* **179**, 2963-2968.

32. Butler, S.L. and Falke, J.J. (1996) Effects of protein stabilizing agents on thermal backbone motions: a disulfide trapping study. *Biochemistry* **35**, 10595-10600.

33. Careaga, C.L. and Falke, J.J. (1992) Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor--detection by disulfide trapping. *J.Mol.Biol.* **226**, 1219-1235.

34. Careaga, C.L., Sutherland, J., Sabeti, J. and Falke, J.J. (1995) Large-amplitude twisting motions of and interdomain hinge--a disulfide trapping study of the galactose-glucose binding protein. *Biochemistry* **34**, 3048-3055.

35. Chaiken, I.M., Freedman, M.H., Lyerla, J.R. and Cohen, J.S. (1973) Preparation and studies of ^{19}F-labeled and enriched ^{13}C-labeled semisynthetic ribonucleas S' analogues. *J.Biol.Chem.* **248**, 884-891.

36. Chang, C. (1996) The ethylene signal transduction pathway in *Arabidopsis*: an emerging paradigm. *Trends Biochem.Sci.* **21**, 129-133.

37. Chen, X. and Koshland, D.E.,Jr. (1997) Probing the structure of the cytoplasmic domain of the aspartate receptor by targeted disulfide cross-linking. *Biochemistry* **36**, 11858-11864.

38. Chen, X.M. and Koshland, D.E.,Jr. (1995) The N-terminal cytoplasmic tail of the aspartate receptor is not essential in signal transduction of bacterial chemotaxis. *J.Biol.Chem.* **270**, 24038-24042.

39. Chervitz, S.A. and Falke, J.J. (1995) Lock-on and -off disulfides identify the transmembrane signaling helix of the aspartate receptor. *J.Biol.Chem.* **270**, 24043-24053.

40. Chervitz, S.A. and Falke, J.J. (1996) Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. *Proc.Natl.Acad.Sci.USA* **93**, 2545-2550.

41. Chervitz, S.A., Lin, C.M. and Falke, J.J. (1995) Transmembrane signaling by the aspartate receptor: engineered disulfides reveal static regions of the subunit interface. *Biochemistry* **34**, 9722-9733.

42. Clarke, S. and Koshland, D.E.,Jr. (1979) Membrane receptors for aspartate and serine in bacterial chemotaxis. *J.Biol.Chem.* **254**, 9695-9702.

43. Cochran, A.G. and Kim, P.S. (1996) Imitation of *Escherichia coli* aspartate receptor signaling in engineered dimers of the cytoplasmic domain. *Science* **271**, 1113-1116.

44. Cohen, C. and Parry, D.A.D. (1990) α-helical coiled coils and bundles: how to design an α-helical protein. *Proteins* **7**, 1-15.

45. Creighton, T.E. (1993) *Proteins: structures and molecular properties*. Ed. 2. W.H. Freeman and Company, New York.

46. Czerlinski, G.H. and Schreck, G. (1964) Chemical relaxation of the reaction of malate dehydrogenase with reduced nicotinamide adenine dinucleotide determined by fluorescence detection. *Biochemistry* **3**, 89-100.

47. Czerlinski, G.H. and Schreck, G. (1964) Fluorescence detection of chemical relaxation of lactate dehydrogenase with reduced nicotinamide adenine dinucleotide. *J.Biol.Chem.* **239**, 913.

48. Dahl, M.K., Boos, W. and Manson, M.D. (1989) Evolution of chemotactic-signal transducers in enteric bacteria. *J.Bact.* **171**, 2361-2371.

49. Danielson, M.A., Bass, R.B. and Falke, J.J. (1997) Cysteine and disulfide scannning reveals a regulatory α-helix in the cytoplasmic domain of the aspartate receptor. *J.Biol.Chem.* (In Press)

50. Danielson, M.A., Biemann, H.-P., Koshland, D.E.,Jr. and Falke, J.J. (1994) Attractant- and disulfide-induced conformational changes in the ligand-binding domain of the chemotaxis aspartate receptor: a F-19 NMR study. *Biochemistry* **33**, 6100-6109.

51. Danielson, M.A. and Falke, J.J. (1996) Bacterial Chemotaxis Proteins: Fluorine-19 NMR. In *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D.M. and Harris, R.K., eds) p. 855-861, Wiley, Chichester.

52. Danielson, M.A. and Falke, J.J. (1996) Use of F-19 NMR to probe protein structure and conformational changes. *Annu.Rev.Biophys.Biomol.Struct.* **25**, 163-195.

53. Darzins, A. (1994) Characterization of a Pseudomonas aeruginosa gene cluster involved in pilus biosynthesis and twitching motility: sequence similarity to the chemotaxis proteins of enterics and the gliding bacterium Myxococcus xanthus. *Mol.Microbiol.* **11**, 137-153.

54. Deckers, H.M. and Voordouw, G. (1994) Identification of a large family of genes for putative chemoreceptor proteins in an ordered library of the *Desulfovibrio vulgaris* Hildenborough genome. *J.Bact.* **176**, 351-358.

55. Dolla, A., Fu, R., Brumlik, M.J. and Voordouw, G. (1992) Nucleotide sequence of dcrA, a *Desulfovibrio vulgaris* Hildenborough chemoreceptor gene, and its expression in *Escherichia coli*. *J.Bacteriol.* **174**, 1726-1733.

56. Drake, S.K., Bourret, R.B., Luck, L.A., Simon, M.I. and Falke, J.J. (1993) Activation of the phosphosignaling protein CheY. *J.Biol.Chem.* **268**, 13081-13088.

57. Ecker, J.R. (1995) The ethylene signal transduction pathway in plants. *Science* **268**, 667-675.

58. Everiss, K.D., Hughes, K.J., Kovach, M.E. and Peterson, K.M. (1994) The Vibrio cholerae acfB colonization determinant encodes an inner membrane protein that is related to a family of signal-transducing proteins. *Infect.Immun.* **62**, 3289-3298.

59. Falke, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A. and Danielson, M.A. (1997) The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu.Rev.Cell.Dev.Biol.* **13**, 457-512.(In Press)

60. Falke, J.J., Blair, D.F., Silhavy, T.J. and Schmitt, R. (1995) BLAST 1995: international conference on bacterial locomotion and signal transduction. *Mol.Microbiol.* **16**, 1037-1050.

61. Falke, J.J., Dernburg, A.F., Sternburg, D.A., Zalkin, N., Milligan, D.L. and Koshland, D.E.,Jr. (1988) Structure of a bacterial sensory receptor: a site-directed sulfhydryl study. *J.Biol.Chem.* **263**, 14850-14858.

62. Falke, J.J. and Koshland, D.E.,Jr. (1987) Global flexibility in a sensory receptor: A site directed cross linking approach. *Science* **237**, 1596-1600.

63. Falke, J.J., Luck, L.A. and Scherrer, J. (1992) ^{19}F nuclear magnetic resonance studies of aqueous and transmembrane receptors. *Biophys.J.* **62**, 82-86.

64. Fersht, A. (1985) *Enzyme structure and mechanism.* W.H. Freeman & Co., New York, N.Y..

65. Foster, D.L., Mowbray, S.L., Jap, B.K. and Koshland, D.E.,Jr. (1985) Purification and characterization of the aspartate chemoreceptor. *J.Biol.Chem.* **260**, 11706-11710.

66. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A. and Perret, X. (1997) Molecular basis of symbiosis between *Rhizobium* and *legumes. Nature* **387**, 394-401.

67. Frillingos, S., Sun, J., Gonzalez, A. and Kaback, H.R. (1997) Cysteine-scanning mutagenesis of helix II and flanking hydrophilic domains in the lactose permease of *Escherichia coli. Biochemistry* **36**, 269-273.

68. Gammon, K.L., Smallcombe, S.H. and Richards, J.H. (1972) Magnetic resonance studies of protein-small molecule interactions. Binding of N-trifluoroacetyl-D-(and L-)-*p*-fluorophenylalanine to chymotrypsin. *J.Am.Chem.Soc.* **94**, 4573-4580.

69. Gardina, P.J., Bormans, A.F., Hawkins, M.A., Meeker, J.W. and Manson, M.D. (1997) Maltose-binding protein interacts simultaneously and asymmetrically with both subunits of the Tar chemoreceptor. *Mol.Microbiol.* **23**, in press.

70. Gardina, P.J. and Manson, M.D. (1996) Attractant signaling by an aspartate chemoreceptor dimer with a single cytoplasmic domain. *Science* **274**, 425-426.

71. Gegner, J.A., Graham, D.R., Roth, A.F. and Dahlquist, F.W. (1992) Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **18**, 975-982.

72. Gerig, J.T. (1989) Fluorine nuclear magnetic resonance of fluorinated ligands. *Methods Enzymol.* **177**, 3-23.

73. Gerig, J.T. (1994) Fluorine NMR of proteins. *Prog.Nucl.Magn.Reson.Spectrosc.* **26**, 293-370.

74. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data [published erratum appears in Anal Biochem 1990 Sep;189(2):283]. *Anal.Biochem.* **182**, 319-326.

75. Gonzalez, L.J., Brown, R.A., Richardson, D. and Alber, T. (1996) Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat.Struct.Biol.* **3**, 1002-1000.

76. Gonzalez, L.J., Woolfson, D.N. and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat.Struct.Biol.* **3**, 1011-1010.

77. Greck, M., Platzer, J., Sourjik, V. and Schmitt, R. (1995) Analysis of a chemotaxis operon in Rhizobium meliloti. *Mol.Microbiol.* **15**, 989-1000.

78. Greene, S.R. and Stamm, L.V. (1997) . *Direct Submission*

79. Gregory, D.H. and Gerig, J.T. (1991) Structural effects of fluorine substitution in proteins. *J.Comp.Chem.* **12**, 180-185.

80. Hagman, K.E., Porcella, S.F., Popova, T.G. and Norgard, M.V. (1997) Evidence for a methyl-accepting chemotaxis protein gene (mcp1) that encodes a putative sensory transducer in virulent *Treponema pallidum*. *Infect.Immun.* **65**, 1701-1709.

81. Hanlon, D.W. and Ordal, G.W. (1994) Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in Bacillus subtilis. *J.Biol.Chem.* **269**, 14038-14046.

82. Hanlon, D.W., Rosario, M.M., Ordal, G.W., Venema, G. and Van Sinderen, D. (1994) Identification of TlpC, a novel 62 kDa MCP-like protein from Bacillus subtilis. *Microbiology.* **140**, 1847-1854.

83. Harkey, C.W., Everiss, K.D. and Peterson, K.M. (1994) The Vibrio cholerae toxin-coregulated-pilus gene tcpI encodes a homolog of methyl-accepting chemotaxis proteins. *Infect.Immun.* **62**, 2669-2678.

84. Haugland, R.P. (1996) *Handbook of Fluorescent Probes and Research Chemicals.* Ed. 6. Molecular Probes, Inc., Eugene, OR.

85. Hazlewood, G.P., Davidson, K., Laurie, J.I., Huskisson, N.S. and Gilbert, H.J. (1993) Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. *J.Gen.Microbiol.* **139**, 307-316.

86. Hess, J.F., Oosawa, K., Kaplan, N. and Simon, M.I. (1988) Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **53**, 79-87.

87. Hoch, J.A. and Silhavy, T.J. (1995) *Two-Component Signal Transduction.* American Society for Microbiology, Washington D.C..

88. Hocking, D., Rothel, L., Doidge, C., Radcliff, F., Lee, A. and Webb, E. (1996) . *Direct Submission*

89. Hofmann, K. and Stoffel, W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol.Chem.Hoppe-Seyler* **347**, 166.

90. Hughson, A.G. and Hazelbauer, G.L. (1996) Detecting the conformational change of transmembrane signaling in a bacterial chemoreceptor by measuring effects on disulfide cross-linking *in vivo*. *Proc.Natl.Acad.Sci.USA* **93**, 11546-11551.

91. Hughson, A.G., Lee, G.F. and Hazelbauer, G.L. (1997) Analysis of protein structure in intact cells: crosslinking in vivo between introduced cysteines in the transmembrane domain of a bacterial chemoreceptor. *Protein.Sci.* **6**, 315-322.

92. Hull, W.E. and Sykes, B.D. (1976) Fluorine-19 nuclear magnetic resonance study of fluorotyrosine alkaline phosphatase: the influence of zinc on protein structure and a conformational change induced by phosphate binding. *Biochemistry* **15**, 1535-1546.

93. Igo, M.M., Ninfa, A.J., Stock, J.B. and Silhavy, T.J. (1989) Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor. *Genes.Dev.* **3**, 1725-1734.

94. Imae, Y., Oosawa, K., Mizuno, T., Kihara, M. and Macnab, R.M. (1987) Phenol: a complex chemoeffector in bacterial chemotaxis. *J.Bact.* **169**, 371-379.

95. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109-136.

96. Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. and Tabata, S. (1995) Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.* **2**, 153-166.

97. Kehry, M.R. and Dahlquist, F.W. (1982) Adaptation in bacterial chemotaxis: CheB-dependent modification permits additional methylation of sensory transducer proteins. *Cell* **29**, 761-772.

98. Kehry, M.R. and Dahlquist, F.W. (1982) The methyl-accepting chemotaxis proteins of *Escherichia coli*. Identification of the multiple methylation sites on methyl-accepting chemotaxis protein. *J.Biol.Chem.* **257**, 10378-10386.

99. Kehry, M.R., Doak, T.G. and Dahlquist, F.W. (1985) Sensory adaptation in bacterial chemotaxis: regulation of demethylation. *J.Bact.* **163**, 983-990.

100. Kehry, M.R., Engstrom, F.W., Dahlquist, F.W. and Hazelbauer, G.L. (1983) Multiple covalent modifications of Trg, a sensory transducer of *Escherichia coli*. *J.Biol.Chem.* **258**, 5050-5055.

101. Khan, S., Amoyaw, K., Spudich, J.L., Reid, G.P. and Trentham, D.R. (1992) Bacterial chemoreceptor signaling probed by flash photorelease of a caged serine. *Biophys.J.* **62**, 67-68.

102. Kim, H.-W., Perez, J.A., Ferguson, S.J. and Campbell, I.D. (1990) The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. *FEBS Lett.* **272**, 34-36.

103. Kim, K.-S., Sackett, R.L. and Farrand, S.K. (1996) . *Direct Submission*

104. Kletzin, A. and Adams, M.W. (1996) Molecular and phylogenetic characterization of pyruvate and 2- ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus*

and pyruvate ferredoxin oxidoreductase from Thermotoga maritima. *J.Bacteriol.* **178**, 248-257.

105. Kneller, D.G., Cohen, F.E. and R. Langridge, R. (1990) Improvements in protein secondary structure prediction by an enhanced neural network. *J.Mol.Biol.* **214**, 171-182.

106. Kolodziej, A.F., Tan, T. and Koshland, D.E.,Jr. (1996) Producing positive, negative, and no cooperativity by mutations at a single residue located at the subunit interface in the aspartate receptor of *Salmonella typhimurium*. *Biochemistry* **35**, 14782-14792.

107. Koshland, D.E.,Jr. (1977) A response regulator model in a simple bacterial sensory system. *Science* **196**, 1055-1063.

108. Krah, M., Marwan, W., Vermeglio, A. and Oesterhelt, D. (1994) Phototaxis of *Halobacterium salinarium* requires a signalling complex of sensory rhodopsin I and its methyl-accepting transducer HtrI. *EMBO J.* **13**, 2150-2155.

109. Krikos, A., Conley, M.P., Boyd, A., Berg, H.C. and Simon, M.I. (1985) Chimeric chemosensory transducers of *Escherichia coli*. *Proc.Natl.Acad.Sci.USA* **82**, 1326-1330.

110. Krikos, A., Mutoh, N., Boyd, A. and Simon, M.I. (1983) Sensory transducers of E. coli are composed of discrete structural and functional domains. *Cell* **33**, 615-622.

111. Kunkel, T.A., Bebenek, K. and McClary, J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125-139.

112. Kuramitsu, H.K. and Kataoka, M. (1995) . *Direct Submission*

113. Kuroda, A., Kumano, T., Taguchi, K., Nikata, T., Kato, J. and Ohtake, H. (1995) Molecular cloning and characterization of a chemotactic transducer gene in *Pseudomonas aeruginosa*. *J.Bacteriol.* **177**, 7019-7025.

114. Le Moual, H. and Koshland, D.E.,Jr. (1996) Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. *J.Mol.Biol.* **261**, 568-585.

115. Lee, G.F., Burrows, G.G., Lebert, M.R., Dutton, D.P. and Hazelbauer, G.L. (1994) Deducing the organization of a transmembrane domain by disulfide cross-linking - the bacterial chemoreceptor Trg. *J.Biol.Chem.* **269**, 29920-29927.

116. Lee, G.F., Dutton, D.P. and Hazelbauer, G.L. (1995) Identification of functionally important helical faces in transmembrane segments by scanning mutagenesis. *Proc.Natl.Acad.Sci.U.S.A.* **92**, 5416-5420.

117. Lee, G.F., Lebert, M.R., Lilly, A.A. and Hazelbauer, G.L. (1995) Transmembrane signaling characterized in bacterial chemoreceptors by using sulfhydryl cross-linking *in vivo*. *Proc.Natl.Acad.Sci.USA* **92**, 3391-3395.

118. Li, E., Quian, S.-J., Nader, L., Young, N.-C.C., d'Avignon, A. and et al., (1989) Nuclear magnetic resonance studies of 6-fluorotryptophan-substituted rat cellular retinol binding protein II produced in *Escherichia coli*. *J.Biol.Chem.* **264**, 17041-17048.

119. Li, J., Li, G. and Weis, R.M. (1997) The serine chemoreceptor from *Escherichia coli* is methylated through an inter-dimer process. *Biochemistry* **36**, 11851-11857.

120. Li, J.Y., Swanson, R.V., Simon, M.I. and Weis, R.M. (1995) The response regulators CheB and CheY exhibit competitive binding to the kinase CheA. *Biochemistry* **34**, 14626-14636.

121. Lin, L.N., Li, J.Y., Brandts, J.F. and Weis, R.M. (1994) The serine receptor of bacterial chemotaxis exhibits half-site saturation for serine binding. *Biochemistry* **33**, 6564-6570.

122. Liu, J. and Parkinson, J.S. (1991) Genetic evidence for interaction between the CheW and Tsr proteins during chemoreceptor signaling by *Escherichia coli*. *J.Bacteriol.* **173**, 4941-4951.

123. Liu, J.D. and Parkinson, J.S. (1989) Role of CheW protein in coupling membrane receptors to the intracellular signaling system of bacterial chemotaxis. *Proc.Natl.Acad.Sci.USA* **86**, 8703-8707.

124. Long, D.G. and Weis, R.M. (1992) Oligomerization of the cytoplasmic fragment from the aspartate receptor of *Escherichia coli*. *Biochemistry* **31**, 9904-9911.

125. Loomis, W.F., Shaulsky, G. and Wang, N. (1997) Histidine kinases in signal transduction pathways of eukaryotes. *J.Cell Sci.* **110**, 1141-1145.

126. Lu, P., Jarema, M., Mosser, K. and Daniel, W.E. (1976) Lac repressor: 3-fluorotyrosine substitution for nuclear magnetic resonance studies. *Proc.Natl.Acad.Sci.USA* **73**, 3471-3475.

127. Luck, L.A. and Falke, J.J. (1991) Open conformation of a substrate-binding cleft--F-19 NMR-studies of cleft angle in the D-galactose chemosesory receptor. *Biochemistry* **30**, 6484-6490.

128. Luck, L.A. and Falke, J.J. (1991) F-19 NMR studies of the D-galactose chemosensory receptor 1. Sugar binding yields a global structural change. *Biochemistry* **30**, 4248-4256.

129. Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem.Sci.* **21**, 375-382.

130. Lupas, A. and Stock, J.B. (1989) Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. *J.Biol.Chem.* **264**, 17337-17342.

131. Lynch, B.A. and Koshland, D.E.,Jr. (1991) Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A.* **88**, 10402-10406.

132. Macnab, R.M. (1996) Flagella and Motility. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhardt, F.C., ed) p. 123-145, ASM Press, Washington D.C..

133. Maddock, J.R., Alley, M.R. and Shapiro, L. (1993) Polarized cells, polar actions. *J.Bacteriol.* **175**, 7125-7129.

134. Maddock, J.R. and Shapiro, L. (1993) Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717-1723.

135. Manson, M.D., Armitage, J., Hoch, J.A. and Macnab, R.M. (1997) Bacterial locomotion and signal transduction. *J.Bact.* (In Press)

136. Manson, M.D. and Kossmann, M. (1986) Mutations in *tar* supress defects in maltose chemotaxis caused by specific *malE* mutants. *J.Bacteriol.* **165**, 34-40.

137. Maruyama, I.N., Mikawa, Y.G. and Maruyama, H.I. (1995) A model for transmembrane signaling by the aspartate receptor based a random cassette mutagenesis and site-directed disulfides. *J.Mol.Biol.* **253**, 530-546.

138. McBride, M.J., Weinberg, R.A. and Zusman, D.R. (1989) "Frizzy" aggregation genes of the gliding bacterium *Myxococcus xanthus* show sequence similarities to the chemotaxis genes of enteric bacteria. *Proc.Natl.Acad.Sci.U.S.A.* **86**, 424-428.

139. McLachlan, A.D. and Stewart, M. (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J.Mol.Biol.* **98**, 293-304.

140. Mesibov, R. and Adler, J. (1972) Chemotaxis toward amino acids in *Escherichia coli*. *J.Bacteriol.* **112**, 315-326.

141. Michotey, V., Toussaint, B., Richaud, P. and Vignais, P.M. (1996) Characterisation of the mcpA and mcpB genes capable of encoding methyl-accepting type chemoreceptors in *Rhodobacter capsulatus*. *Gene* **170**, 73-76.

142. Milburn, M.V., Prive, G.G., Milligan, D.L., Scott, W.G., Yeh, J., Jancarik, J., Koshland, D.E.,Jr. and Kim, S.-H. (1991) Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science* **254**, 1342-1347.

143. Milligan, D.L. and Koshland, D.E.,Jr. (1988) Site-directed cross-linking. Establishment of the dimeric structure of the aspartate receptor of bacterial chemotaxis. *J.Biol.Chem.* **263**, 6268-6275.

144. Milligan, D.L. and Koshland, D.E.,Jr. (1991) Intrasubunit signal transduction by the aspartate chemoreceptor. *Science* **254**, 1651-1654.

145. Milligan, D.L. and Koshland, D.E.,Jr. (1993) Purification and characterization of the periplasmic domain of the aspartate receptor. *J.Biol.Chem.* **268**, 19991-19997.

146. Mirmira, R.G. and Tager, H.S. (1991) Disposition of the phenylalanine-B-25 sidechain during insulin-receptor and insulin-insulin interactions. *Biochemistry* **30**, 8222-8229.

147. Moe, G.R, Bollag, E. and Koshland, D.E.,Jr. (1989) Transmembrane signaling by a chimera of the *Escherichia coli* aspartate receptor and the human insulin receptor. *Proc.Natl.Acad.Sci.USA* **86**, 5683-5687.

148. Morimoto, B.H. and Koshland, D.E.,Jr. (1991) Short-term and long-term memory in single cells. *FASEB J.* **5**, 2061-2067.

149. Mowbray, S.L., Foster, D.L. and Koshland, D.E.,Jr. (1985) Proteolytic fragments identified with domains of the aspartate receptor. *J.Biol.Chem.* **260**, 11711-11718.

150. Mowbray, S.L. and Koshland, D.E.,Jr. (1987) Additive and independent responses in a single receptor: aspartate and maltose stimuli on the Tar protein. *Cell* **50**, 171-180.

151. Murray-Rust, P., Stallings, W.C., Monti, C.T., Preston, R.K. and Glusker, J.P. (1983) Intermolecular interactions of the C-F bond: the crystallographic environment of fluorinated carboxylic acids and related structures. *J.Am.Chem.Soc.* **105**, 3206-3214.

152. Nara, T., Kawagishi, I., Nishiyama, S., Homma, M. and Imae, Y. (1996) Modulation of the thermosensing profile of the *Escherichia coli* aspartate receptor tar by covalent modification of its methyl-accepting sites. *J.Biol.Chem.* **271**, 17932-17936.

153. Ninfa, E.G., Stock, A., Mowbray, S.L. and Stock, J. (1991) Reconstitution of the bacterial chemotaxis signal transduction system from purified components. *J.Biol.Chem.* **266**, 9764-9770.

154. Noback, M.A., Terpstra, P., Holsappel, S., Venema, G. and Bron, S. (1997) . *Direct Submission*

155. Nowlin, D.M., Bollinger, J. and Hazelbauer, G.L. (1987) Sites of covalent modification in Trg, a sensory transducer of *Escherichia coli. J.Biol.Chem.* **262**, 6039-6045.

156. Oosawa, K. and Simon, M.I. (1986) Analysis of mutations in the transmembrane region of the aspartate chemoreceptor in *Escherichia coli. Proc.Natl.Acad.Sci.USA* **83**, 6930-6934.

157. Ota, I.M. and Varshavsky, A. (1993) A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566-569.

158. Pakula, A.A. and Simon, M.I. (1992) Determination of transmembrane protein structure by disulfide cross-linking: the *Escherichia coli* Tar receptor. *Proc.Natl.Acad.Sci.USA* **89**, 4144-4148.

159. Paliakasis, C.D. and Kokkinidis, M. (1992) Relationships between sequence and structure for the four-α-helix bundle tertiary motif in proteins. *Protein Eng.* **5**, 739-748.

160. Parkinson, J.S. (1993) Signal transduction schemes of bacteria. *Cell* **73**, 857-871.

161. Parkinson, J.S. and Blair, D.F. (1993) Does *E. coli* have a nose? *Science* **259**, 1701-1702.

162. Parkinson, J.S. and Kofoid, E.C. (1992) Communication modules in bacterial signaling proteins. *Annu.Rev.Genet.* **26**, 71-112.

163. Pauling, L. (1960) *The nature of the chemical bond.* Ed. 3. Cornell Univ. Press, Ithaca, N.Y..

164. Peersen, O.B., Pratt, E.A., Truong, H.-T.N., Ho, C. and Rule, G. (1990) Site-specific incorporation of 5-fluorotryptophan as a probe of the structure and function of the membrane-bound D-lactate dehydrogenase of *Escherichia coli:* a 19F nuclear magnetic resonance study. *Biochemistry* **29**, 3256-3262.

165. Post, J.F., Cottam, P.F., Simplaceanu, V. and Ho, C. (1984) Fluorine-19 nuclear magnetic resonance study of 5-fluorotryptophan-labeled histidine-binding protein J of *Salmonella typhimurium*. *J.Mol.Biol.* **179**, 729-743.

166. Pratt, E.A. and Ho, C. (1975) Incorporation of fluorotryptophans into proteins of *Escherichia coli*. *Biochemistry* **14**, 3035-3040.

167. Putkey, J.A., Dotson, D.G. and Mouawad, P. (1993) Formation of inter-and intramolecular disulfide bonds can activate cardiac troponin C. *J.Biol.Chem.* **268**, 6827-6830.

168. Quast, U., Engel, J., Heumann, H., Krause, G. and Steffen, E. (1974) Kinetics of the interaction of bovine pancreatic trypsin inhibitor (Kunitz) with alpha-chymotrypsin. *Biochemistry* **13**, 2512-2520.

169. Rice, M.S. and Dahlquist, F.W. (1991) Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. *J.Biol.Chem.* **266**, 9746-9753.

170. Richards, F.M. (1977) Areas, volumes, packing and protein structure. *Annu.Rev.Biophys.Bioeng.* **6**, 151-176.

171. Rudolph, J., Nordmann, B., Storch, K.F., Gruenberg, H., Rodewald, K. and Oesterhelt, D. (1996) A family of halobacterial transducer proteins. *FEMS.Microbiol.Lett.* **139**, 161-168.

172. Rule, G.S., Pratt, E.A., Simplaceanu, V. and Ho, C. (1987) Nuclear magnetic resonance and molecular genetic studies of the membrane-bound D-lactate dehydrogenase of *Escherichia coli*. *Biochemistry* **26**, 549-556.

173. Russell, C.B., Stewart, R.C. and Dahlquist, F.W. (1989) Control of transducer methylation levels in Escherichia coli: investigation of components essential for modulation of methylation and demethylation reactions. *J.Bacteriol.* **171**, 3609-3618.

174. Russo, A.F. and Koshland, D.E.,Jr. (1983) Separation of signal transduction and adaptation functions of the aspartate receptor in bacterial sensing. *Science* **220**, 1016-1020.

175. Sahin-Toth, M. and Kaback, H.R. (1993) Cysteine scanning mutagenesis of putative transmembrane helice-IX and helix-X in the lactose permease of *Escherichia coli*. *Protein Sci.* **2**, 1024-1033.

176. Saxild, H. (1996) . *Direct Submission*

177. Schuster, S.C., Swanson, R.V., Alex, L.A., Bourret, R.B. and Simon, M.I. (1993) Assembly and function of a quaternary signal transduction complex monitered by surface plasmon resonance. *Nature* **365**, 343-347.

178. Scott, W.G. and Stoddard, B.L. (1994) Transmembrane signalling and the aspartate receptor. *Structure* **2**, 877-887.

179. Seeley, S.K., Weis, R.M. and Thompson, L.K. (1996) The cytoplasmic fragment of the aspartate receptor displays globally dynamic behavior. *Biochemistry* **35**, 5199-5206.

180. Segall, J.E., Manson, M.D. and Berg, H.C. (1982) Signal processing times in bacterial chemotaxis. *Nature* **296**, 855-857.

181. Seidel, R., Scharf, B., Gautel, M., Kleine, K., Oesterhelt, D. and Engelhard, M. (1994) . *Direct Submission*

182. Seidel, R., Scharf, B., Gautel, M., Kleine, K., Oesterhelt, D. and Engelhard, M. (1995) The primary structure of sensory rhodopsin II: a member of an additional retinal protein subgroup is coexpressed with its transducer, the halobacterial transducer of rhodopsin II. *Proc.Natl.Acad.Sci.U.S.A.* **92**, 3036-3040.

183. Sekiguchi, J. (1997) . *Direct Submission*

184. Shapiro, L. (1993) Protein localization and asymmetry in the bacterial cell. *Cell* **73**, 841-855.

185. Shapiro, M.J., Chakrabarti, I. and Koshland, D.E.,Jr. (1995) Contributions made by individual methylation sites of the *Escherichia coli* aspartate receptor to chemotactic behavior. *Proc.Natl.Acad.Sci.USA* **92**, 1053-1056.

186. Shibui, H., Hamamoto, T., Yohda, M. and Kagawa, Y. (1997) The stabilizing residues and the functional domains in the hyperthermophilic V-ATPase of *Desulfurococcus*. *Biochem.Biophys.Res.Comm.* (In Press)

187. Springer, W.R. and Koshland, D.E.,Jr. (1977) Identification of a protein methyltransferase as the CheR gene product in the bacterial sensing system. *Proc.Natl.Acad.Sci.USA* **74**, 533-537.

188. Spudich, E.N., Hasselbacher, C.A. and Spudich, J.L. (1988) Methyl-accepting protein associated with bacterial sensory rhodopsin I. *J.Bact.* **170**, 4280-4285.

189. Stock, A., Chen, T., Welsh, D. and Stock, J. (1988) CheA protein--a central regulator of bacterial chemotaxis--belongs to a family of proteins that control gene expression in response to changing environmental conditions. *Proc.Natl.Acad.Sci.USA* **85**, 1403-1407.

190. Stock, A., Mottonen, J., Chen, T. and Stock, J. (1987) Identification of a possible nucleotide binding site in CheW, a protein required for sensory transduction in bacterial chemotaxis. *J.Biol.Chem.* **262**, 535-537.

191. Stock, A.M. and Mowbray, S.L. (1995) Bacterial chemotaxis--a field in motion. *Curr.Op.Struct.Biol.* **5**, 744-751.

192. Stock, J.B., Lukat, G.S. and Stock, A.M. (1991) Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. *Annu.Rev.Biophys.Biophys.Chem.* **20**, 109-136.

193. Stock, J.B. and Surette, M.G. (1996) Chemotaxis. In *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F.C., ed) p. 1103-1129, ASM Press, Washington, D.C..

194. Stoscheck, C.M. (1990) Quantitation of protein. *Methods Enzymol.* **182**, 50-68.

195. Surette, M. and Stock, J.B. (1996) Role of alpha-helical coiled-coil interactions in receptor dimerization, signaling, and adaptation during bacterial chemotaxis. *J.Biol.Chem.* **271**, 17966-17973.

196. Surette, M.G., Levit, M., Liu, Y., Lukat, G., Ninfa, E.G., Ninfa, A. and Stock, J.B. (1996) Dimerization is required for the activity of the protein histidine kinase CheA that mediates signal-transduction in bacterial chemotaxis. *J.Biol.Chem.* **271**, 939-945.

197. Swanson, R.V., Alex, L.A and Simon, M.I. (1994) Histidine and aspartate phosphorylation: two component systems and the limits of homology. *Trends Biochem.Sci.* **19**, 485-490.

198. Swanson, R.V. and Simon, M.I. (1994) Signal transduction: bringing the eukaryotes up to speed. *Current Biology* **4**, 234-237.

199. Sykes, B.D. and Hull, W.E. (1978) Fluorine nuclear magnetic resonance studies of proteins. *Methods Enzymol.* **49**, 270-295.

200. Tanaka, H., Osaka, F., Ohashi, M., Shiraki, M. and Munekata, E. (1986) Substance-P analogues containing para-fluoro-L-phenylalanine. *Chem.Lett.* 391-394.

201. Tatsuno, I., Homma, M., Oosawa, K. and Kawagishi, I. (1996) Signaling by the *Escherichia coli* aspartate chemoreceptor Tar with a single cytoplasmic domain per dimer. *Science* **274**, 423-425.

202. Terwilliger, T.C. and Koshland, D.E.,Jr. (1984) Sites of methyl esterification and deamidation on the aspartate receptor involved in chemotaxis. *J.Biol.Chem.* **259**, 7719-7725.

203. Terwilliger, T.C., Wang, J.Y. and Koshland, D.E.,Jr. (1986) Surface structure recognized for covalent modification of the aspartate receptor in chemotaxis. *Proc.Natl.Acad.Sci.USA* **83**, 6707-6710.

204. Tomb, et al., (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539-547.

205. Tso, W.-W. and Adler, J. (1974) Negative chemotaxis in *Escherichia coli*. *J.Bacteriol.* **118**, 560-576.

206. Utsumi, R., Brissette, R.E., Rampersaud, A., Forst, S.A., Oosawa, K. and Inouye, M. (1989) Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate. *Science* **245**, 1246-1249.

207. van Iwaarden, P., Pastore, J.C., Konings, W.N. and Kaback, H.R. (1991) Construction of a functional lactose permease devoid of cysteine residues. *Biochemistry* **30**, 9595-9600.

208. Vine, W.H., Brueckner, D.A., Needleman, P. and Marshall, G.R. (1973) Synthesis, biological activity, and ^{19}F nuclear magnetic resonance spectra of angiotensin II analogs containing fluorine. *Biochemistry* **12**, 1630-1637.

209. Vogel, H.J. and Bonner, D.M. (1956) Acetylornithinase of *Escherichia coli*: Partial purification and some properties. *J.Biol.Chem.* **218**, 97-106.

210. Wagner, G. and Wütrich, K. (1986) Observation of internal motility of proteins by nuclear magnetic resonance in solution. *Methods Enzymol.* **131**, 307-326.

211. Wang, H. and Matsumura, P. (1996) Characterization of the $CheA_S$ / CheZ complex - a specific interaction resulting in enhanced dephosphorylating activity on CheY-phosphate. *Mol.Microbiol.* **19**, 695-703.

212. Wang, H. and Matsumura, P. (1997) Phosphorylating and dephosphorylating protein complexes in bacterial chemotaxis. *J.Bacteriol.* **179**, 287-289.

213. Wang, J.X., Balazs, Y.S. and Thompson, L.K. (1997) Solid-state REDOR NMR distance measurements at the ligand site of a bacterial chemotaxis membrane receptor. *Biochemistry* **36**, 1699-1703.

214. Ward, M.J., Harrison, D.M., Ebner, M.J. and Armitage, J.P. (1995) Identification of a methyl-accepting chemotaxis protein in Rhodobacter sphaeroides. *Mol.Microbiol.* **18**, 115-121.

215. Weis, R.M. and Koshland, D.E.,Jr. (1988) Reversible receptor methylation is essential for normal chemotaxis of *Escherichia coli* in gradients of aspartic acid. *Proc.Natl.Acad.Sci.USA* **85**, 83-87.

216. Wells, J.A. (1991) Systematic mutational analyses of protein-protein interfaces. *Meth.Enzymol.* **202**, 390-411.

217. Wilson, M.L. and Dahlquist, F.W. (1993) Membrane protein conformational change dependent on the hydrophobic environment. *Biochemistry* **24**, 1920-1928.

218. Wu, J. and Kaback, H.R. (1997) Helix proximity and ligand-induced conformational changes in the lactose permease of *Escherichia coli* determined by site-directed chemical crosslinking. *J.Mol.Biol.* **270**, 285-293.

219. Wu, J.G., Li, J.Y., Li, G.Y., Long, D.G. and Weis, R.M. (1996) The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry* **35**, 4984-4993.

220. Wu, J.R., Long, D.G. and Weis, R.M. (1995) Reversible dissociation and unfolding of the *Escherichia coli* aspartate receptor cytoplasmic fragment. *Biochemistry* **34**, 3056-3065.

221. Wurgler-Murphy, S.M. and Saito, H. (1997) Two-component signal transducers and MAPK cascades. *Trends.Biochem.Sci.* **22**, 172-176.

222. Yamamoto, K. and Imae, Y. (1993) Cloning and characterization of the *Salmonella typhimurium* specific chemoreceptor Tcp for taxis to citrate and from phenol. *Proc.Natl.Acad.Sci.USA* **90**, 217-221.

223. Yang, Y., Park, H. and Inouye, M. (1993) Ligand binding induces an asymmetrical transmembrane signal through a receptor dimer. *J.Mol.Biol.* **232**, 493-498.

224. Yao, V.J. and Spudich, J.L. (1992) Primary structure of an archaebacterial transducer, a methyl-accepting protein associated with sensory rhodopsin I. *Proc.Natl.Acad.Sci.USA* **89**, 11915-11919.

225. Yeh, J., Biemann, H.P., Pandit, J., Koshland, D.E.,Jr. and Kim, S.-H. (1993) The three-dimensional structure of the ligand-binding domain of a wild-type bacterial chemotaxis receptor. *J.Biol.Chem.* **268**, 9787-9792.

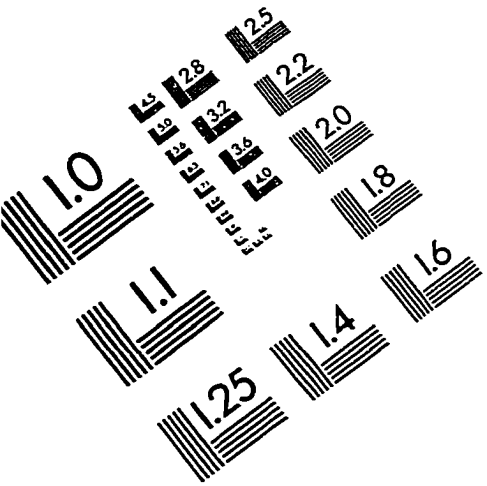
226. Yeh, J.I., Biemann, H.P., Prive, G.G., Pandit, J., Koshland, D.E.,Jr. and Kim, S.H. (1996) High resolution structures of the ligand binding domain of the wild type bacterial aspartate receptor. *J.Mol.Biol.* **262**, 186-201.

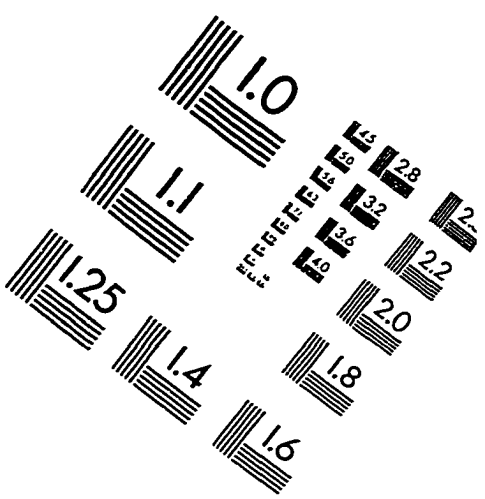
227. Yost, C.K. and Hynes, M.F. (1997) . *Direct Submission*

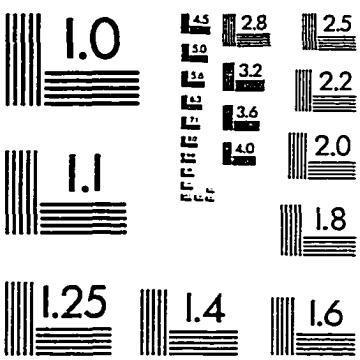
228. Zhang, W., Brooun, A., McCandless, J., Banda, P. and Alam, M. (1996) Signal transduction in the archaeon *Halobacterium salinarium* is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins. *Proc.Natl.Acad.Sci.U.S.A.* **93**, 4649-4654.

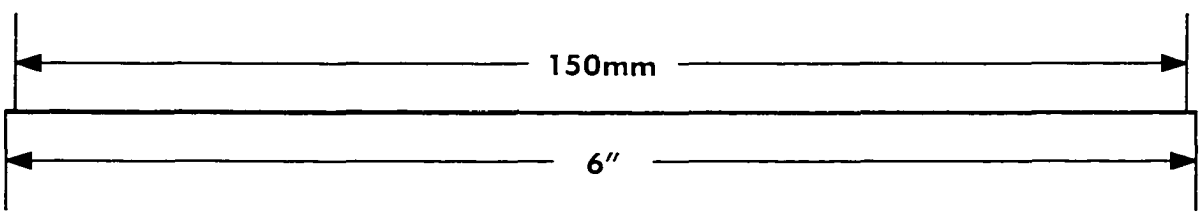
229. Zhang, W., Brooun, A., Mueller, M.M. and Alam, M. (1996) The primary structures of the Archaeon Halobacterium salinarium blue light receptor sensory rhodopsin II and its transducer, a methyl- accepting protein. *Proc.Natl.Acad.Sci.U.S.A.* **93**, 8230-8235.

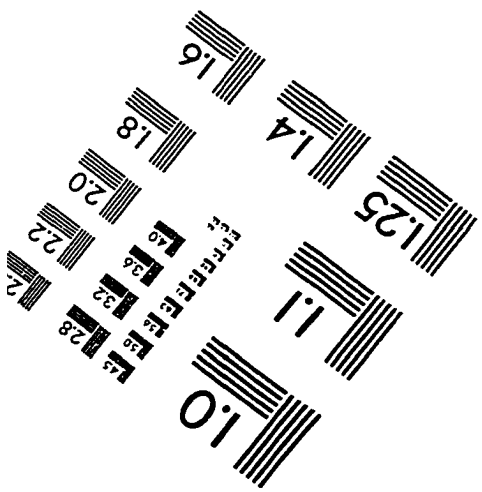
IMAGE EVALUATION TEST TARGET (QA-3)

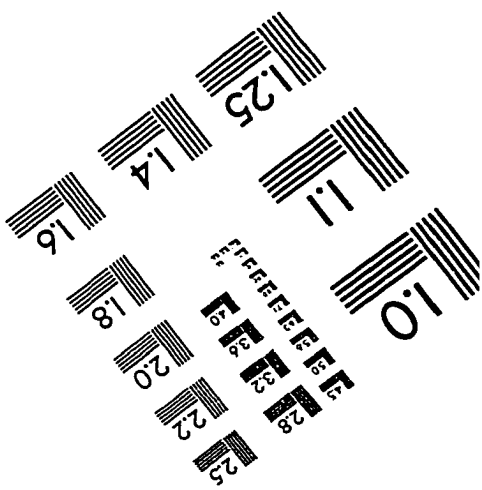












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

DEVELOPING A LITERATE IDENTITY

THROUGH

PARTICIPATION ON A HIGH SCHOOL NEWSPAPER

by

Kathryn A.H. Davinroy
B.A., The Pennsylvania State University, 1984
M.A., The Pennsylvania State University, 1989

A thesis submitted to the Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
School of Education
1998

This thesis entitled:
Developing a Literate Identity through Participation on a High School Newspaper
written by Kathryn A.H. Davinroy
has been approved for the School of Education

Dr. Shelby A. Wolf,
Chair of Committee

Dr. Hilda Borko,
Committee member

Dr. Margaret Eisenhart,
Committee member

Dr. Steven Guberman,
Committee member

Dr. William McGinley,
Committee member

Dr. Cynthia White,
Committee member

Date 3/31/98

The final copy of this thesis has been examined by the signators,
and we find that both the content and the form meet acceptable presentation
standards of scholarly work in the above mentioned discipline.

Davinroy, Kathryn A.H. (Ph.D., Education)

Developing a Literate Identity through Participation on a High School Newspaper

Thesis directed by Dr. Shelby A. Wolf

This study explores the developing literate identities of 23 high school students producing their school newspaper, sharing the goal of "getting out the news through the power of words."

Qualitative methods of data collection and analysis were guided by current theories of genre and a sociohistorical approach to learning and teaching. Genre, long constrained by traditional modes and surface regularities, is understood as the contextualized whole of recurring rhetorical situations, textual production, and social practices that take place regularly within a community. Work in this classroom is Activity as defined by Activity Theorists and is thus a system of tasks and actions performed to achieve a collective goal.

Initially, students viewed writing as a teacher-directed, school-based assignment. Over time, their perceptions of literacy, especially writing, evolved to embrace writing for personal voice and public efficacy. Students moved along membership trajectories from newcomer to veteran and came to see themselves as writing agents for social action. Importantly they became members of a team whose success depended upon dedication and responsible performance.

As students took on the group's norms and ideologies, they simultaneously developed writerly identities reflecting their social and classroom statuses. Mini-case studies of three students represent the voices of individuals illustrating the range of participation possibilities. Variety and flexibility of roles allowed development of personal ways of writing, crucial for personal satisfaction and group success.

Students performed valuable social--arguably critical--work through their written words. They reflected on changing their minds or the minds of their peers

about issues like AIDS, teen pregnancy, and censorship. Risk taking in writing elicited tangible changes in the school.

This research suggests ways for rethinking writing pedagogy, curriculum, and teacher education. When writing is real--with real consequences, audience, and issues--it leads youth to engage with social worlds outside the teacher-student dyad. A genre approach to writing pedagogy places the student community at the center, encourages ownership and responsibility for literate acts, and puts skills and strategies in a real context.

To *Mom and Dad*

Acknowledgments:

The linear march of words across the page do no justice to the depth of gratitude I feel toward those who have been supportive and patient with me and my work.

I wish to acknowledge the enthusiastic participation of the teacher facilitator and the student staff who warmly welcomed me into their production activities. Their words and actions continually invigorated and rejuvenated me throughout the entire process of this research. I have come to care deeply for their work and words.

I can never thank Shelby Wolf for her kind wisdom and generous intelligence. For it has been through her unflagging energy and spirit that I have completed this work.

And I thank you, Tom.

CONTENTS

CHAPTER		PAGE
1.	INTRODUCTION	1
	Myself Becoming a Reader of World and Word	6
	Genre: From Traditional Forms to Social Action	7
	Genre as the Bridge from Pretend to the Real World	13
2.	CONCEPTUAL FRAMEWORK: CONSTRUCTING AN ACTIVITY-BASED THEORY OF GENRE	18
	Language for Social Action	18
	Genre as Social Action: Creating an Activity-Based Theory of Genre	21
	Doing, Saying, and Writing: An Activity-Based Theory of Genre	30
	Situating the Writer within Activity	32
	Activity, Actions, and Tasks: Connecting Principles of Activity with Genre	35
	Activity: The first level of analysis	36
	Actions: The second level of analysis	37
	Tasks: The third level of analysis	39
	Learning Genre: Participating in Activity as Part of a Writerly Identity	43
	Conclusion	47
3.	DESCRIPTION OF METHODS	48
	Where? Who? and What? Description of Site, Participants, and the Newspaper	50
	Where? A Description of the Site	52
	Who? A Description of the Students and Teacher	53
	Mr. Tillinger: Teacher and Managing Editor	58
	What? Description of the Newspaper, the Zephyr	63
	Data Collection	68
	Data Analysis	73
	Stage One Analysis: General Impressions and Broad Categories	77
	Literacy	79
	Audience	81
	Individuality and Identity	82
	School	83
	Professionalism	84
	Teamwork and Images of Family	85
	Stage Two Analysis: Applying Activity Theory	87
	Analyzing the First Level of Activity Settings: The Activity	89
	Analyzing the Second Level of Activity Settings: The Actions	92
	Analyzing the Third Level of Activity Settings: The Task	100
	Stage Three Analysis: Microanalysis of Individual Participation	102
	Stage Four Analysis: Rhetorical and Textual Analysis of Writing	107
	Analysis of Surface Features	108
	Rhetorical Features of Text	110
	Social Features	111
	Preview of Findings	112
4.	DEVELOPING A WRITERLY IDENTITY	116
	Setting the Theoretical Stage: Linking Genre and Identity	119
	Peopling the Stage: What is the Activity Like?	123
	The Roles: Playing and Being	127

The Activity: Ways of Doing on the Zephyr..... 129
The Actions: Enacting the Production Process..... 131
The Action of Writing..... 132
The Action of Team building..... 146
At the Junction of Writers, Social Practices, and Texts..... 156

5. CASES OF WRITERS WRITING..... 161
Mark: The Crusader..... 168
The "Etiquette" of Writing for the Zephyr..... 171
Social World of Mark's Writing and Membership..... 184
Cass: The Writer..... 187
Developing Her Writerly Self..... 191
Cass' Social World on the Zephyr..... 204
Dean: The Man Who Would Be Editor..... 207
Dean as Writer..... 211
Dean's Social World on the Zephyr: Holding Court..... 219
Mark, Cass, and Dean: Individuals Writing Their Ways into Different Social Worlds..... 223
Looking Across Three Experiences..... 225
Group Rules and Self Expression..... 230

6. TAKING SOCIAL ACTION THROUGH WRITING..... 237
Understanding the "Critical" in Critical Literacy..... 238
Taking Social Action through Writing..... 243
Self as Site: Places in the Heart and in the World..... 247
Changing my mind as social action..... 249
"Arm yourself with knowledge"..... 255
Reaffirming strongly held beliefs..... 259
Others as Site for Change: Changing Other People's Minds..... 261
Negotiating audience..... 263
School as site for change: Working on the community..... 266
Exploring the Conditions for Critical Literacy Practice..... 269
Power: Teacher in charge..... 269
Students at the helm: Power handed over..... 272
Active participation..... 274
The risky business of writing..... 277
Public recognition for the written word..... 280
On Using Literacy Critically..... 282

7. TAKING A GENRE APPROACH TO WRITING..... 286
What It Means to Take a "Genre Approach" to Teaching Writing..... 287
Community in the Classroom: Overlapping Formal and Informal Social Practices..... 290
A Teach Within and Across Boundaries..... 295
Student Writers and the Liminal World..... 298
Genre as Textual Practices..... 302
Caveats for Future Work..... 304
In the End..... 306

TABLES

TABLE		PAGE
3.1	Profile of Students on the Zephyr Staff	54
3.2	Descriptive Characteristics of the Zephyr	65
3.3	Description of Data Collection	70
3.4	Linking Research Questions with Types of Data	74
3.5	Coding Categories for Motivation for Activity from the Wider Social World	91
3.6	Categories for Relational Analysis	96
3.7	Categories of Stances Taken Toward Others	97
3.8	Linking Actions to Their Constitutive Tasks	101
3.9	Surface Features and Descriptors of Textual Analysis	109
3.10	Categories for Rhetorical Features of Language	110
3.11	Connecting Writing with Social Self	113
5.1	Comparison of Newspaper Subgenres for Mark	171
5.2	Topic Categories for Straight News and Features for Mark	173
5.3	Comparison of Newspaper Subgenres for Cass	192
5.4	Rhetorical Features of Cass' Feature Writing	197
5.5	Comparison of Topic-Types for Dean	213
5.6	Comparison of Subgenres within Sports Stories for Dean	214
5.7	Comparison of Subgenres Across Mark, Cass, and Dean	229

FIGURES

FIGURE		PAGE
3.1	Proportion of high, middle, and low ability students	55
3.2	Peer-group memberships for Zephyr staff	56
3.3	Proportions of girls and boys according to grade	57
3.4	Definitions and terminology used on the Zephyr	69
3.5	Outline of stages of analysis	76
3.6	Key linkages between social practices and illustrative verbal markers	94
3.7	Example of coded portion of data using stance-taking and relational analyses	99
3.8	An example of rhetorical analysis of student writing	112
5.1	Distribution of Dean's news stories over the course of the year	212

CHAPTER 1
INTRODUCTION

LIVING WITH AIDS

By SERA NIEMAN
Zephyr News Staff

Jane Camary walks into the room with a Peabody's coffee and her keys in one hand and in her other hand a book on AIDS education. She says she needs the coffee because she's a little phased from her all-day drive from New Mexico; she is still awake. Camary is a 31 year old CPA, post professional chef, white, heterosexual, middle-class, ex-cross country running, HIV positive woman.

"About two years ago I went to get an HIV test because I had convinced all my gay friends to get tested, and they were saying, 'Why don't you do it too?' so I did and when I went in to get the results the doctor guy looked at my chart and told me I was HIV negative. While we were talking about that, I looked at my chart and saw that the box checked HIV positive was the one that was marked. I asked him about that and he was completely shocked. He had never given an HIV positive diagnosis to a woman." Camary explained.

Women targeted

Right now, women are the fasted growing group of HIV infected people in the United States. For this reason, Camary targets most of her speaking to women. "Women are eighteen times more likely to become infected if their partner is positive. Also, women are not taught to negotiate condom use," she says.

Camary's appearance is not one of the dying AIDS victims who has physically wasted away. She is taller than average and has maintained target-level weight for her height. This normal, heterosexual woman is infected with a virtually 100% terminal disease.

Camary says, "It's such an ugly disease. I take about 52 pills a day because when you're HIV positive, your body doesn't absorb minerals."

In addition to having to physically alter her lifestyle, Camary has also had to deal with massive emotional changes.

"Now it's very difficult to relate to my peers. I'm trying to make peace with it, but while people in my age group are talking about kids and retirement funds, I'm thinking how do I want my memorial service, will I be cremated, do I want to go to the Hospital or Homecare?"

...

When asked what it is like to live with HIV, Camary gets quiet and then says, "You can't know what it's like to live with HIV. When people ask me what is it like, I just want to grab these people and shake them and say, 'Don't be like me, grab a condom!' This is an ugly way to live and die. You get cancer in your stomach, and you can't eat. You get KS and get lesions all over your body. You get pneumonia and suffocate. You can get dementia and have Alzheimer's at 31. All this is completely avoidable."

...

This feature article written by Sera covered all the column space on page 11 in the October 6, 1996 issue of the Zephyr, the student-run newspaper serving Langston High. The excerpt here captures the writing style, the attention to detail, and, importantly, the passion that Sera developed as a staff writer for the paper. Dressed in plain brown pants, red corduroy shirt, and a turned-around ball cap over her long, dark blond hair, Sera's writing stood out more than her appearance. She looked at her

hands in her lap and told me, "After that article, I wanted to shake people too. Just like Jane. This was a turning point for me; I knew I'd be writing stories to challenge people to think" (4/12/96). Sera wrote this social commentary feature article after several weeks of investigative reporting. For her the things she remembered most about writing this article included learning the "science of AIDS," understanding the terrible loneliness of living with AIDS, her struggle with "writing it just so, so that adults wouldn't worry about 'children' being exposed to this topic in school," and her subsequent decision to volunteer at the county AIDS project (4/12/96). Sera's involvement with this social issue, and later with others, was a result of her decision to write about a socially relevant topic. Her choice to write about AIDS reflects her developing beliefs in the power of words to evoke change in her community. Her choice also reverberates with a wider social commitment to freedom of speech and press, to curiosity and learning, and to development of compassion and community.

Beliefs about the power and importance of writing, beliefs like Sera's, train through our collective psyche and are made manifest through our literature, through the plethora of newspapers (surviving still in paper form), and even our governing constitution which recognizes the critical necessity for language to range over the national landscape: recording, entertaining, persuading. To understand these social qualities of language is to become literate. To use literacy as a tool for personal and social efficacy is to develop a writerly identity. The study reported here is about my understanding of one group of young people as they wrote their newspaper pieces for social action. In this introduction to my research on the building of literate identities in a high school newsroom, I will not condense the findings of my project; it is not an overview. Rather, it attempts to set the stage for the questions that guided the research. The central question was:

How does the activity of a high school newspaper create and recreate social action for youth through written and spoken genres?

More specifically, I asked:

- How does the wider society contribute to the creation of a high school newspaper, and how does the newspaper in turn affect the wider society?
- What actions constitute the creation of a high school newspaper, and how do people conform to and reform these actions in becoming members of the classroom newspaper?
- What are the writing processes used by individuals for personal expression, and how are these tasks and operations used to create and recreate genres for social action through public expression?

In these questions my relationships with students, teachers, writing, and myself were sustained during the nearly two years of data collection, analysis, and writing. The following pages connect these questions about youth writers to a wider social world and make sense of that connection through my own beliefs and experiences with writing, as well as through the words of those I have read and admired for their thinking about writing.

For teachers and researchers of literacy, connecting people, especially students, to literate lives is a critical part of their work. Where that connecting point is--through reading, writing, listening, or speaking--varies across research agendas and teaching assignments. However, most agree that literacy is part of engagement, participation, and involvement with a variety of communities that must eventually transcend the in-school experience. For many in the writing research community, writing isn't writing unless it has personal and social meaning, much as reading isn't reading without comprehension or interpretation. Thus a person becomes "writerly" (Elbow, 1986, p. xii) when s/he has envisioned and practiced some of the possibilities for writing that lay outside the institutionalized writing that often occurs in school. Peter Elbow, long an advocate of authentic and meaningful writing instruction, explained that the detachment students tend to feel toward their school writing experiences too easily leads to

alienation from a powerful tool for social involvement whether it is economic, political, material, or personal. Writing must be alive with purpose and consequence in order to matter to the writer. Students need to experience the possibilities of being writerly before deciding to become writerly.

In some ways, Sera's personal story of coming to writing--of becoming writerly--through her involvement with the Zephyr, is similar, if less dramatic than, the stories of recent literary newspaper writer protagonists. Quoyle (Proulx, 1993), Ishmael (Guterson, 1994), Ward (Dexter, 1995), and others represent their respective artist-writers' vision of the power of journalism. In all these novels, the possibilities that grow out of writing are a central feature in the protagonists' connection and disconnection to community and self. These characters experience life-changing events related to their decisions to write the world they read and live in. Their experiences, though fictional, reflect a larger social preoccupation with the importance and power of written words and tell the story of coming to writerly selves amidst social and personal challenges.

Because of the prevalence of these literary news writing figures in literature, I draw upon their fictional voices from time-to-time to open chapters, to make points about writing, and to help draw conclusions about the writing experiences that are possible for some youth. This crossing of genres in my dissertation, an academic text, is, in part, my attempt to link literary and research worlds, for there is much to be learned about a nation's beliefs and values about literacy by looking to its literature (Willinsky, 1990). Genre crossing is also a critical feature of my research study, and it seems fitting to blur the margins of dissertations just as the participants in my study blurred their margins of journalistic writing and student behavior.

My text, however, is not fiction. But also it is not Truth. This text is as much my story as it is the story of Sera and the 22 other writers for the high school newspaper, the Zephyr. The genre crossing that these youth engaged in is akin to the

genre crossing in which I have engaged. Moving from literary critic to educational researcher, from writing teacher to writer of dissertation has loosened the margins of research and practice, of fact and fiction, freeing me to see what the students writing the Zephyr think about as they produce their newspaper. The processes of thinking and doing are never completely transparent to any researcher. Thus, as I constructed this text to be loyal, accountable, and true to the events I saw, my beliefs and experiences were a first filter, writing a second, and their responses to my findings a third. For as words move linearly across the page, they create a false sense of certainty and order. That order is in contrast to the simultaneity of writing, thinking, doing, editing, talking, problem-solving, confronting, arguing, explaining that took place every moment of my data collection. The image of words will always be flatter than reality. I am acutely aware that the words I have written can capture only a part of the world I was privy to during this study. Thus, this dissertation says *some* things, not all things, about how the students writing their high school newspaper understood writing and their involvement with a literate world.

If youth are going to develop personally meaningful understandings of their place in a literate world--understandings of writing and literacy as tools with potential--they need images of writing and writers that are not impoverished by narrow visions of schools and schooling. Too often in school, youth are exposed to the extremes of the real writer's world. On the one hand, students read excellent literature and sometimes not-so-excellent textbooks. They see the slick pages, the finished look, the absence of error. They see those products as the work of "someone else," someone who is inspired, gifted, talented, but different from themselves. On the other hand, students see their own written pieces that bear no resemblance to the work of the texts they read. It's scratchy, blotchy, marked up. It doesn't flow; it's awkward, vague, undeveloped, boring. Their work is not that of a writer, but of some kid saying what the teacher wants him to, dreading the sure-to-come red marks on the page. The gap between the

smudged, graded paper and the crisp justified margins of the book is too large to brook, too empty to fill. For many youth, the image of a writer is decidedly "not me." In my two years with the writers of the Zephyr, I shared company with youth who had a "this is me" image of themselves as writers.

Myself becoming a Reader of World and Word

In seeking to understand how youth can come to see themselves as writers, I became a reader of the Zephyr writing-world through my observations and interactions with the staff. Daily, I saw young people energetically involved in the time consuming and challenging activity of producing a newspaper, a vision that contrasted sharply with our current media's focus on the negative actions of youth. I also became a reader of the published word. As I learned more about the meaning of writing and school to these students, I began to connect what I was seeing with research literature.

As I read researchers' words about the meaning of communities of practice and of writers, I saw ways the two worlds, the Zephyr world and the literacy research world, resonated with one another. To these connections, I paid close attention. I consumed Peter Elbow's (1986) reflections on teaching writing--what works, what does not--and Paulo Freire's (1985) work with peasants in a critical literacy paradigm as they worked toward emancipation through reading their worlds and writing their words. I attended to activity theorists (Leontiev, 1981) who tangled with visions of learning through activity. I bent my ear to Toni Morrison (1992; 1994), Helene Cixous (1991) and a host of other published authors who assured me that with something to say, my mind would be filled and fingers fly. I read the literary newspapermen who pay occasional visits to this study affirming a social consciousness about writing and its place in this society.

This dissertation explores the work and words of the young people writing the Zephyr through the lens of my own experience and beliefs in literacy and especially

writing as part of success in various communities. I have formalized my lens by searching the literacy research literature for concepts that help shape my beliefs. Thus, I constructed an academic conceptual scaffold that represents deeper beliefs and values that I hold. I sought the words of others to help me describe the shape, size, scope, and meaning of my personal beliefs about writing. This "clear" version of the concepts is a trick of written language. We know that thinking and life itself aren't nearly so clean, neat, and packaged. That said, sharp lines of definition, here, serve the purpose of setting the margins, for now--of flattening for a moment the complexity of living writers.

Genre: From Traditional Forms to Social Action

Margin setting on paper can be thought of as a metaphor for rules and boundaries around ideas as well as communities. The concept of genre has long set the margins of meaning and form for writing. Recently, however, genre has been the subject of extensive debate in literacy circles. Historically, it was seen in terms of set regularities and specific textual features. Indeed, genres have long been defined in formalist and structuralist terms based upon surface regularities in texts. This structuralist understanding of genres has helped dichotomize literacy processes into reading and writing, text and context, and forms and functions. Separating these aspects from one another disallows a view of the processes as "constitutive of one another" (Kamberelis, 1995, p. 118). Genres conceived purely in formalist ways are "static and normalizing structures that constrain individuals and determine the outcomes of communicative events" (Kamberelis, 1995, p. 118). These static and normalizing features of genres have supported a foundation for writing courses where teachers identify textual forms and teach writing in an algorithmic fashion. But this conceptualization of genre doesn't allow for the ways in which writers work genres and even blur genre boundaries as they write for particular audiences.

Because genre expectations grow from co-participation of readers and writers, experiments in genre blurring are common. Journalism, for example, has certain historical markers that signal the reader to expect specific kinds of styles, themes, and contents. With its heavy emphasis on objective language and "just the facts ma'am" approach, young journalists are schooled to believe they can capture a truth, a moment. Newspapers are expected to clearly demark the shifts between facts and opinion. Yet, as Tannen (1989) notes, journalistic writing that uses details for "'vivid' writing (rather than accurate, clear, or informative writing) seems out of place in reference to journalism....[yet] it is use of language in which it is the language itself that counts most...and [is] illustrative of effective use of details" (p. 160). Thus, writers and readers come to understand that effective writing often means "blurring genres"--as in using literary language to make information vivid or using journalistic language to capture the details of events in order to interpret meanings.

If learning these ways of effective writing involves the ability to recognize genre conventions that may be appropriate in new and unusual places, then novice writers need multiple opportunities to try these cross-genre experiments in authentic contexts. In one novice-writing setting--school--these opportunities are, in general, limited to teacher generated assignments rather than centered on student-relevant contexts. But, there is in some schools at least one context for students to experiment and practice writing for a real world audience--the school newspaper. At Langston High, the Zephyr is not a "junior" or novice version of a New York Times. Rather, it is a collectively written document which blurs genres of traditional journalism with other written genres as a way for students to engage in social action.

Indeed, forms found in the Zephyr reflect the blurring of journalism genres with pop art genres. As members of the self-described "MTV generation" (4/15/96), the students have altered the graphic look of the paper peppering its pages with bold designs and logos. Experiments in form take place in answer to the function of the

writing: traditional "sections" of a newspaper are renamed for the purposes of the student audience. Thus, the "Style" section of the Washington Post becomes the "Paris Connect" section in the Zephyr. While these forms may not conform to traditional notions of newspapering, they do carry the stamp of past journalists and the forging of a style, form, and content appropriate for the messages moving from this newsroom to their readers. The genre of the school newspaper carries with it the indelible mark of its readers and writers, of the social action deployed through literacy.

Theorizing about genres as "social action"--language action taken within a particular context and meaning system--is supported by current literacy community enthusiasm for Bakhtin's (1986) ideas of intertextuality and "speech genres." Contemporary discussions of genre fit a sociohistorical view of language and meaning as socially constructed and evolutionary (Vygotsky, 1978). Language evolves into patterns of use that serve members of a community. It is collectively and pragmatically defined by the action that is to be taken by the language-user in particular contexts (Rorty, 1989). Thus, genres, as the regularities evident in speech and writing (Bakhtin, 1986), are defined as social actions to be taken in "recurrent rhetorical situations" (Miller, 1984).

These recurrent rhetorical situations are the social contexts in which the regularities in spoken and written language are directly linked to the social action taken by orators and writers. They can be as diverse as the conversations around a dinner table and as rigid as a business writing textbook. Genres, both spoken and written, are intrinsically linked to these situations as they serve symbolically to represent the message of a collective view of appropriate language use. In traditional rhetorical theory, genres referred to specific opportunities for public oratory, opportunities guided by rules handed down through years of tradition. Modern thinking about rhetorical situations refers to the traditional opportunities to communicate persuasively with a real audience but take a broader view of genres to reflect patterns in language that emerge in

many situations rather than a few pre-determined and formal opportunities. A recurring rhetorical situation would be one where the context in which communication occurs is repeated, perhaps cyclically, and in which the orators or writers are attempting to achieve some sort of persuasive purpose. A student newspaper such as the Zephyr would be one such "recurring rhetorical situation" or activity in which written genres are created and recreated to achieve some social action, or serve persuasive purposes. Through the student newspaper genre, words emerging from this recurring rhetorical situation reflect the worlds of the students.

Research on student writing suggests that the words of young people describing their school-based worlds often reflect an adult focus on correctness and evaluation (Applebee, 1989; Wolf, 1994). Studies show that students are extremely adept at figuring out what teachers want and at writing for a teacher-evaluator. However, the words of the writers on the Zephyr reflect a different emphasis. The words are enthusiastic: "I stay late"; "I love seeing my name in print"; "it's for the students not the teacher." The words are urgent: "How can they give money to prisons not schools?" "Why are we piercing our bodies?" The words are compassionate: "I remember the first time the Editor worked over my piece"; "Let's do the front page for John [a fellow student who died in a car accident]." These writers look for words to inform, inflame, incense--the right words to describe their worlds. Hearing the students' voices can be a great education for educators, as Jesse noted, "I decided, and now I know it was right. I wanted to write where it would go somewhere and people would see it....Here, it's more real life than anything else because you have a real product that you come out with. It's not for pretend" (Jenne, 4/15/96).

Writing "not for pretend," writing for the real world, is critical to the social meaning that the writers on the Zephyr are making around their writing and classroom activity. According to the students, writing that goes to a real reader is important writing: it gives voice and meaning to personal experiences that can be shared with

others in a wider community. At the same time, a writer, writing with the voice of being a member of a community, represents aspects of that community through generic regularities specific to the communication context. Thus, a writer who is a member of a writing community speaks through the collective voice of the community as well as through a personal voice. It is the personal voice used within the collective context that has potential power to shift and recreate the genres established through a group's history. And it is through the genres of the community that individuals can perhaps work to make larger changes in their social lives.

That literacy, especially writing, can create change in the social world and unlock the door to powerful futures for students is virtually an educational commonplace.[1] However, "language and transformative agency" (for example, Freire, 1985; Freire & Macedo, 1987; Giroux, 1988; Lankshear & McLaren, 1993) make for compelling theory, but is such possibility realized in the literacy learning in schools? In considering language as one key to personal and public efficacy, we need to know what genres provide students with opportunities for the social action called for by critical literacy researchers (e.g., Freire, 1973; Lankshear & McLaren, 1993). Understanding these uses of language for students as relationships among texts, genres, and social practices (Kamberelis, 1995) might help educators see ways that literacy instruction is or can be more "personally meaningful, culturally relevant, and motivating" (McGinley & Kamberelis, 1992).

Deciding whether acquiring genres through implicit or contextualized ways (genre as social action) or "learning" language through explicit, "universal" rule-binding lessons (genre as static and normalizing) has been the source for heated debate among English educators (Applebee, 1989; Freedman, 1993). But chasing this issue as a question of one or the other can lead research on genre acquisition to chicken-and-

[1] In Graff's (1986) historical research on literacy and literate practices in the US, he takes to task the general assumption that literacy is empowering socially and economically. He calls for research that could help support such claims.

egg itself into forgetting who counts most: the learners and users of language genres. Students. Teachers get no closer to helping students write more effectively when research factions advocate diametrically opposed theories of pedagogy and learning. For example, while some research claims that explicit teaching is "ineffective" or "unnecessary" for learning to use language (e.g., Freedman, 1993); other research reports that prescribing memorized lock-step incremental rules is the best method for language learning (e.g., Weaver & Kintsch, 1991). More productively, a combination of perspectives on genre contributes to an understanding of literacy that can lead toward developing a literate identity. For in understanding why some writing works in particular contexts, we may come to understand how to teach about genres appropriate for communicating within those contexts. And conversely, by understanding how and when writers blur and deploy their genres, we can better understand why and how certain forms do their social "work" within certain communities.

Learning about the social work of writing and the social action taken through writing for these youth requires understanding the overall activity and the constituent actions of community members that enable them to journey toward becoming effective writers for the Zephyr. How genre gets done and how activity gets done are mutually constitutive and intimately linked to understanding the literate identities that students adopt and adapt as they become members of the school news staff.

Indeed, the generic concept of the recurring rhetorical situations can also be considered an activity in the sense used by the Soviet Activity Theorists (Leontiev, 1981). As Berkenkotter & Huckin (1993) point out,

> Miller's (1984) major contribution to the discussion of genre was to take the notion of genre as recurrent response to a rhetorical situation... [which] extricates the concept of genre from its moorings in Aristotelian and literary classification systems and relocates it in a[n]...understanding of communicative behaviors (p. 480).

This un-mooring of genre from its past of being "steeped in formalist and structuralist epistemologies" (Kamberelis, 1995, p. 118) allows for a rethinking of genre as a

socially constructed activity in which participants engage in the production of literacy. "An activity based theory of genre knowledge would therefore locate our learning of academic genres...in the situated activities of a practitioner in training" (Berkenkotter & Huckin, 1993, p. 487). Because activities are described as systems of patterned actions and operations derived from participants, social practices and goals, as well as reflections of wider social motives, they may also be defined as "recurring rhetorical situations."

The contextualization of the recurring regularities of text and the actions of writers link it to knowledge and language as socially constructed. Individuals developing a literate identity call upon and create understandings of genre to make sense of their communications; receivers of messages likewise call upon their knowledge of genre to make sense of the message. Using genres is a part of being a member of a literate community. It is the communication-product both created and constrained by the communicating group as well as the receiving group. Indeed, Daiute (1993) notes, "In relation to literacy, a culture becomes concrete in its genres--the types of written texts and oral routines characteristic among its people" (p. 403). We can understand a community, in part, by studying its genres; we can understand genres by studying the communities from which they emerge. And in considering genres within their "recurring rhetorical situations," we can learn something about the meaning, learning, and uses of writing for members of that community.

Genre as the Bridge from Pretend to the Real World

In this study of students writing a high school newspaper, understanding how this group of students creates a collectively written product out of individually written stories offers insight into the process by which students appropriate and reshape generic language for social action. A critical question is how does the activity of a high

school newspaper create and recreate social action--and, therefore, meaning--for youth through written and spoken genres?

There are many ways to approach answers to this central question. Some researchers might note that this is a group of students creating a "novice" or beginner newspaper and might compare their writing and speech with similar exchanges occurring in the newsroom of a professional daily paper. One aspect of their social action would be to demonstrate membership in a mature journalism community. Elaine Chin's (1993) study of student journalists in an introductory college level news writing course considered this movement of student writing toward a professional news story. She looked for markers of language, such as use of compelling leads, development of story through objective language, and the appropriation of journalistic talk when writers discussed their stories with each other. Writing for Chin's study was assigned and graded by the course professor. Formal features of mature journalism were held out by the professor as the target; evaluation was carried out by the professor. Comparing early student writing with professional news writing enables some measurement of student journalists as they move along a learning trajectory (Lave & Wenger, 1991) from student to real-world journalist and suggest an apprenticeship model where students are learning to act like journalists though a context that would best be described as university classroom.

Other researchers might look at this scene and think of it as a classroom writing event. Students are perceived as novice-writers engaged in producing written documents for a teacher to grade. Comparing student writing with teacher-standards is typical classroom procedure for many engaged in teaching and learning to write (Applebee, 1989; Flower, 1989; McLaughlin, Talbert & Bascia, 1990). Researchers explore how well students internalize or absorb the teachers' expectations and standards by defining "excellence" before students' produce their work. Learning to write, from this psychologized perspective, is considered in terms of the individual and how s/he

composes his/her own text. The primary focus is on the individual as a developing writer in the teacher's eyes, and the research lens turns toward how students use the particular classroom occasion to become a better writer in terms set out by the teacher.

However, neither of these models for learning to write seems to take into account what writing *does* in the world for writers, how writers undertake social action with their writing. These models for research which focus on understanding student writers as junior journalists or as junior school-writers provide only a partial view of the complex interrelationships between individual learning and the social nature of many language acts--a concern voiced broadly in educational research and specifically in the literacy and composition research communities (Carter, 1990; Fingert, 1994; Flower, 1989; Fulkerson, 1990; Lensmire, 1995). These concerns usher in yet another perspective for understanding this scene by shifting the analysis from comparing student performance with formally defined genres (like becoming a "journalist" and writing a "good" paper for school) to focusing, instead, on what the students are doing and saying as they engage in the activity of producing their newspaper for their schoolmates. From an activity perspective, students can be seen as engaging in a community of learning (Brown, Ash, Rutherford, Nakagawa, Gordon, & Campione. 1993) for particular purposes and as making decisions about genre based upon their socially constructed understanding of what needs to be said or written and why.

This study begins, then, with a classroom that sits at the junction of the "real world" demands of journalists--where participants rarely conduct explicit lessons on writing--and the traditional "school based" demands of teachers and students--where little real-world, audience-driven literacy activity occurs. Such a context may be authentic in valuable ways for developing writers. And it is in contrast with a body of research (Applebee, 1989; Doyle, 1986; Flower, 1989; Hairston, 1982) which shows teachers and students engaged in a sort of pseudo-real-world writing where teachers assign types of writing and students generally write to hypothetical audiences.

Teachers grade the success of writing based upon their understandings and expectations for the hypothetical audience. While teachers strive to make writing contexts relevant and authentic, in reality, students rarely see real world consequences for their words.

For the Zephyr writers, both demands are operating in ways that shape and create knowledge of writing as a socially constituted activity. Scribner and Cole (1973) noted the tensions between these two arenas for learning: "School represents a specialized set of educational experiences which are discontinuous from those encountered in everyday life....[School] requires and promotes ways of learning and thinking which often run counter to those nurtured in practical daily activities" (p. 555). Research on these "nontraditional" or "hybrid" classrooms is slim (Tonso, 1994), but reveals interesting features of student learning as they participate in work and writing that is destined for someone besides the teacher. While such classrooms have been in existence for some time (e.g., experiential classes, vocational education), usually research on literacy has focused on the "traditional" classroom. Perhaps something can be learned about writing and learning to write from the class that engages in the "real life" writing that students say is so valuable for their writing experiences.

Researchers have called for "close readings" of students' writing worlds. Discovering the meanings of writing to youth and how that meaning is manifest through writing activity will help teachers consider ways to build classrooms for effective writing instruction. Hanna Fingert (1994) points out the shift in literacy research that provides a support for these interests:

> Researchers have come to think of literacy as broader than reading text; we speak of reading the world, of constructing the meanings of symbols and connecting our meanings to our actions. Using terms like *whole language, authentic assessment, and participatory literacy education,* we've come to see literacy as interactive, culturally related, and collaborative. Literacy is fundamentally a social construct, relying on shared meanings for language and symbols. Literacy is complicated. (p. 31)

Writing for self, others, and a wider world are indeed complicated processes. These complications arise as literacy educators consider better ways to align school work with real life literacy acts, and as real life asks that schools prepare students for the demands of a post-industrial workplace (Heath, 1991). By looking at ways that some classrooms can meet these diverse and at times divergent learning goals researchers can help guide others in developing meaningful writing contexts for learners--contexts that are "for real" not "pretend."

CHAPTER 2

CONCEPTUAL FRAMEWORK:

CONSTRUCTING AN ACTIVITY-BASED THEORY OF GENRE

To understand how the activity of a high school newspaper creates and recreates social action for youth through their written and spoken genres, we need a framework that draws upon a theory for learning that is grounded in social interaction. We need, as well, a theory of writing and genre that moves beyond the printed page into the worlds of the writers-in-activity. Recent theoretical work on genre, largely in the fields of rhetoric and composition as well as communications, provides a solid base to build such a framework. In addition, activity theory provides a multi-layered view of learning through collective and shared work, and a view through which genre and other literacy actions can be explored.

When asking how students create social action through the genre of the school newspaper, we must consider a number of relationships: how written genres work as critical reflections and transformations of a wider society; how the interactions of the group members are constructed as they speak and write in generic forms appropriate for their group membership as well as for their readership; and how the writer creates a relationship with his/her writing, through the processes, risks, and creative enterprises s/he undertakes. Understanding genre as social action means understanding these different layers of relationship and the interconnections among them. For to take social action, the individual writer does not operate in isolation, but instead works in carefully choreographed interactions with self, other staff members, and with a wider society of peers and adults.

Language for Social Action

> *"Whether it is obscuring state language or the faux-language of mindless media; whether it is the proud but calcified language of the academy or*

> *the commodity driven language of science; whether it is the malign language of law-without-ethics, or language designed for the estrangement of minorities, hiding its racist plunder in its literary cheek--it must be rejected, altered and exposed....There is and will be rousing language to keep citizens armed and arming; slaughtered and slaughtering in the malls, courthouses, post offices, playgrounds, bedrooms and boulevards; stirring memorializing language to mask the pity and waste of needless death....Language alone protects us from the scariness of things with no names."*
>
> --Toni Morrison,
> Lecture in honor of her Nobel Prize in Literature, Stockholm, 1993

Faux-language. Academic language. Science language. Law language. These genres each have specific formal characteristics as well as social power--power to disenfranchise or call-to-arms, power to destroy or create, power to rend or bind. In Morrison's view, it is the task of humanity to use language for social action and for protection. Yet how will language "protect us from the scariness of things with no names?" For Morrison, protection and power come through learning to use language genres for social action, to make change in the world, to learn both about and through language. She suggests that we use language to hold us to dreams, to join us in communities, and to commit us to social justice. To use language in these ways, then, is indeed to take social action. Thus, knowing *how* language works historically, politically, economically, as well as personally is as important as being affected by language (Freire & Macedo, 1987; Halliday, 1989; Willinsky, 1991). To be empowered through and with language, it isn't enough to use it--only to know *of* language--one must be able to know why and how it works. Rhetorically. In context. With motive and intent. And a person's literate identity is critical to this empowerment.

But how can students come to see themselves as empowered users of language, as writers with social purpose? Critical literacy proponents would say they need a social context--a purpose, audience, and topic--that holds meaning for them (Freire, 1993; Freire & Macedo, 1987; Giroux, 1988; Lankshear & McLaren, 1993). Some teaching and learning theorists would say they need to experience literacy-in-action;

they need to see, to feel the power and the possibility of words, especially their own words and voices, in activity (Rogoff, 1990; Vygotsky, 1987).

Many researchers of language have become interested in the effects of social interaction on learning both about and through language (Fingert, 1994; Heath, 1983; Lankshear & McLaren, 1993; Rosenblatt, 1978; Willinsky, 1991). Understanding language genres within the context of their use in social activity is one way to learn how individuals within particular communities take on as well as change the genres for speaking, listening, reading, and writing (Heath, 1983 & 1991; Leontiev, 1981; Rogoff, 1990; Wertsch, 1981; Wolf & Heath, 1992). Indeed, some argue that when individuals "take on" the genres of their communities, they may be reconstructing their identities as well (Kamberelis, 1995). Because my own argument is designed to answer the question of how the activity of a high school newspaper creates and recreates social action for youth through written and spoken genres, literate *activity* must be the research focus of observation and involvement. Through an activity framework of literacy-in-action I hope to better understand the genre of the school newspaper as critical social action--an activity linking the individual through literacy and a literate voice to self, community, and wider society.

Literacy researchers have begun to make connections between the individual as language learner and the individual as language user, as agent in community (Freire, 1985, 1993; Freire & Macedo, 1987; Giroux, 1988). Explicating these connections can show how language functions to perpetuate, create, and sometimes disturb knowledge and genre stability. It can also show how individuals simultaneously develop personal literate voices as well as voices that speak for a particular community of which the speaker is a member. And as Morrison calls us to use language for social action, so too do educators and emancipators--all those who believe that language can "protect us from the scariness of things with no names." For in naming the world, we can look at, discuss, and change that world (Berthoff, 1987).

This research framework considers concepts of literacy pertinent to understanding the social action taken by students through the genre of student newspaper writing. I will explore these concepts through the lens of Activity Theory with its three-tiered focus on activity, action, and operation. Thus concepts like critical literacy and genre will be linked to the Activity level of analysis in Activity theory, as these concepts directly link the literacy user to a wider social context, a society outside the writer that motivates the actions and operations necessary for the functioning of the group effort to produce a newspaper. Concepts in writing research like rhetorical situation and authenticity are linked to the action level of analysis in Activity Theory, as they reflect the social practices among members of the group as they proceed about their daily actions and interactions of creating news for their school. And finally, concepts like writing process will be linked to the task and operation level of Activity Theory as this reflects a more personal involvement of the individual writer creating a story for the paper through his/her own processes.

Because definitions provide the heart of any research and guide all subsequent uses of the term in that research, I first provide a general overview of recent understandings and work in genre theory and Activity Theory. Then I show how literacy concepts and principles of genre guiding this research can be thought of as "belonging" in the three levels of Activity Theory. This three-layered scaffold will then serve as the infrastructure for the data collection and analyses that are outlined in chapter three.

Genre as Social Action:
Creating an Activity-Based Theory of Genre

Genre is a critical concept when considering the meanings and reception--the social action and intent--of any writer's writing, for genre is wrapped in historical,

political, and creative efforts of individuals and groups. To understand the writing produced by a group of young people who meet regularly to create and produce the particular genre of a school newspaper, we need to understand that genre as an historical expression of that community, as an embodiment of that community's political motives and beliefs, and as a creative experiment in written communication. To aid in the unpacking of the activity of writing, genre theory, then, must move beyond the written product and explore the particular social contexts and conditions in which writing occurs.

Indeed, recent theoretical work on genre has shifted its conceptualization away from traditional structuralist approaches (Aristotle, 1978; Fowler, 1982). The view of genre that has dominated the past thirty or forty years has been rooted squarely in Aristotle's *Poetics* along with other rhetoricians of antiquity, for example, Cicero and Socrates, which "foreground the reception of texts (reading/hearing) and background their production (writing/speaking)" (Kamberelis, 1995, p. 118). Instead, genre now is being conceputalized as "social action taken in recurring rhetorical situations" (Miller, 1984). Such a view radically shifts studies of genre away from a sort of "template approach" in which writers study-to-copy forms created by other writers in other situations to a closer exploration of what writers intend and produce in response to a felt need to communicate. Kamberelis (1995) notes this intimate connection between form and function: "Genres imply not only particular types of texts but also particular processes of producing, distributing, and receiving texts" (p. 116). Inherent in this view is the notion that the social context greatly influences the writer's choices and decisions for risk taking through language genres. Kamberelis (1995) sums up this conceptual shift:

> Most of this [genre] theory and research redirects our attention away from textual forms and toward the people, institutions, purposes, thematic content, rhetorical situations, social contexts, specific text-making practices, and historical trajectories that together constitute genres and are, in turn constituted by genres. This redirection

> reconstitutes genres as loosely defined cultural models consisting of fairly stable constellations of discursive features, thematic content, and rhetorical practices that embody the social histories, norms, and ideologies of the communities or disciplines in which they are commonly used to enact relatively specific rhetorical purposes in fairly recurrent social activities. (p. 119)

The school newspaper is itself a written genre. It is a recurring social activity in which students work to create the relatively stable features and thematic content that reflect and transform their communities. To understand how writing is social action for these students, it will be necessary to learn about their genres, about their social practices. This student newspaper is not an inexpert attempt to be like a Wall Street Journal, nor is it a mere collection of student writing. Instead, it is the social action of a group of students who simultaneously speak through a collective voice, the paper, and with individual voices, in single-authored stories. These simultaneous literacy acts capture the tension of operating as an individual within the constraints of a larger society, and the tension of genre as both prescribed linguistic product and generative social practice.

This tension between stability and dynamism in social practices for language users as well as for spoken and written genres is one of several themes in genre theory research identified by Kamberelis (1995). The other themes he identified are equally important to understanding genres as extending beyond formal and structural textual characteristics. Kamberelis (1995) prefaces his delineation of themes with an important caveat: "discovering or constructing the primary themes in recent...work on genre that cut across vast and diverse literatures is not an easy task; it is also one that is arbitrary" (p. 120). While discovering these themes is probably something more rigorous than arbitrary, it points out the complexity of studying a literacy concept like genre within disciplinary fields with different "world views." Yet, the relationships identified by many theoreticians require more than isolating attention to text, *or* social context, *or* critical action. Instead, a more wholistic view of genre as part of a system of participants, actions, and material products may help educators construct meaningful communities and opportunities for language learners.

Research themes outlined by Kamberelis (1995) provide background to the newer views of genre that guided my research of the Zephyr:

> First, genres, texts, and social practices are understandable only in relation to one another. Second, genres develop historically as both durable and dynamic forms of communicative actions....Third, genres are not simply formal structures but complex configurations of formal features, thematic content..., and social practices of production, distribution, and reception. Fourth, genres develop within and end up "belonging to" fairly specific collectives or social fields. As such, they function as social glue within those communities or fields of practice....Fifth, genres are learned and used as a function of situated practice within multiple embedded activity settings, communities, and social fields. (p. 121)

These themes clarify the "insider" research necessary for understanding the work of writers, especially when that work is explicitly connected to real readers. Bakhtin (1981) argued similarly for contextualized understandings of genres.

> For the speakers of the language themselves, these generic languages and professional jargons are directly intentional...; but outside, that is, for those not participating in the given purview, these languages may be treated as objects, as typifications, as local color. For such outsiders, the intentions permeating these languages become things, limited in their meaning and expression. (p. 289)

Thus an outsider to the community will be limited in his/her understanding of the texts, genres, and social practices of the speakers and writers until s/he spends sufficient time "living" the ways of the community through its language genres. The insiders create, maintain, and use their genres for social action in a wider society. To understand the genres and the ways users take on the generic forms means looking at the language acts as an insider (or as much as is feasible).

The themes outlined by Kamberelis (1995) point to a wide variety of theoretical perspectives and considerations of genre, yet all share a common foundation in defining genre as more than formal features of texts, rather as socially shaped and received practice. Genres, then, are

> human activity structures that shape and are given shape by specific social and cultural fields of discourse production, distribution, and reception....[G]enres are organized according to relatively predictable

> constellations of form/activity at several levels: discourse [textual] form, thematic content, and social practice. (p. 165)

As these themes and theories of genre are sifted down through thinking and rethinking, through theoretical and empirical research, constituent parts of the social action taken by writers through their genres are revealed. In cyclic fashion these constituent parts, or characteristics, provide others with structure for empirical research.

From their long term research on adult writers in disciplinary communities, Berkenkotter and Huckin (1993) propose five constituent features of genre that they have arrived at inductively through their data as well as through wide-ranging reading in many disciplines, including: "structuration theory in sociology, rhetorical studies, interpretive anthropology, ethnomethodology, Bakhtin's theory of speech genres, Vygotsky's theory of ontogenesis, and Russian Activity Theory" (p. 478). From this array of disciplinary theories and from their first-hand experience with writers, they propose five principles that, for them, constitute a "theoretical framework" (p. 478). These five principles are:

> 1. Dynamism: Genres are dynamic rhetorical forms that develop from responses to recurrent situations and serve to stabilize experience and give it coherence and meaning. Genres change over time in response to their users' sociocognitive needs.
>
> 2. Situatedness: Our knowledge of genre is derived from and embedded in our participation in the communicative activities of daily and professional life. As such, genre knowledge is a form of "situated cognition," which continues to develop as we participate in the activities of the culture.
>
> 3. Form and content: Genre knowledge embraces both form and content, including a sense of what content is appropriate to a particular purpose in a particular situation at a particular point in time.
>
> 4. Duality of structure: As we draw on genre rules to engage in professional activities, we *constitute* social structures (in professional, institutional, and organizational contexts) and simultaneously *reproduce* these structures.
>
> 5. Community ownership: Genre conventions signal a discourse community's norms, epistemology, ideology, and social ontology. (p. 478)

These five principles reflect the themes outlined by Kamberelis (1995), and support the co-constitutive relationship among genres, texts, and social practice. This sociocognitive approach to thinking about genre proposed by Berkenkotter and Huckin (1993) provides a framework within which to consider genre as a social action (Miller, 1984).

But where do we find genres as social action in real life rather than in theory? And when we do find them, how do we understand genres as expressions compared not with some expert model, but rather as intentionally crafted expressions of the community from which the genres come? This is a research and methodological problem that emerges when genre theory is no longer bound to the formal and surface characteristics of text but is rather perceived through the lens of sociocognitive glasses. As a number of genre theorists have noted, "To date, very little work on genre in rhetorical studies has been informed by actual case research with insiders" (Berkenkotter & Huckin, 1993, p. 477).

Indeed, recent empirical research on genre that has taken the newer conceptualization of genre as dynamic, situated, constituted of form and content, dually structured, and owned by a community is often "encased" (Geertz, 1983, p. 4) by conventional disciplinary boundaries. For example, many recent published studies of genre consider the acquisition of the more stable features of genres located mainly within academic communities and mainly as perceived in written documents. Here, genres in science and law have received a lot of research attention.

Two well-cited studies of scientific article writing (Bazerman, 1988; Huckin, 1987) show the dynamism of genre through the evolution over time of the science article which is often considered a rigid example of written genre. Close textual analysis showed changes that included "increased foregrounding of experimental results in titles, abstracts, introductions, and section headings, and a devaluation of explanations of procedures and methods" (Huckin, 1987, p. 478). Huckin (1987)

situated these changes in "reception practices" described by scientists in interviews. Scientists noted their increasing reading load and decreasing time allotments for reading, necessitating reading practices that could be described as "skimming." Huckin (1987) notes that the genre of the scientific article is bound to the social context of the scientific writers: "their reading behavior affects their writing strategies" (p. 481). Analysis of written documents, as in these genre research studies, offers proof of the dynamic and social nature of writing. Huckin's interviews with scientists offer some social context for the shifts in form and content; however, this version of social context emphasizes the individual as s/he reads and writes texts rather than in direct verbal and written interaction with others in the disciplinary community.

Chin's (1994) work with journalism students in a college setting found the disjunctions between school activity systems and "real world" activity systems led students to make genre decisions that would aid them in future job searches. The genre decisions signaled the social action, or actually *were* the social actions taken by students when confronted with choices between activity systems. In her research, she found that once students learned what would count in the real world of journalism through their internships, they divested their interests in the teacher-made assignments and re-created the assignments to write "clips" for their portfolios. In other words, they appeared to use these genre choices as a way to begin reconstructing their identities more in line with real world journalism expectations and shifted out of or away from their student identities. Chin's (1994) work shows "that practitioners develop a kind of social consciousness of the organizational environment" (Paradis, Dobrin, & Miller, 1985, p. 193). *Social* consciousness is, thus, connected to ways of thinking and using language that make practitioners members of a community, that enable members to identify with particular work and world communities.

As these studies show, much research on genre has, by and large, concentrated on academic disciplines and on mature writers, often in post-secondary and out-of-

school environments (Swales, 1990). While the authors of these studies argue that the social context of the adult writers guided their acquisition of the genre forms appropriate for action within that community, these contexts are not richly described in ways that show how the verbal and written interactions among members help construct the "ways with words" anchored to the local community or to the wider society. Indeed, the heavy focus on the written documents leaves it unclear as to how the adults in the studies consider themselves part of a discourse community at all. Additionally, these studies reflect attention to individual writers producing individual texts as they "come up" into mature practices (Prior, 1994), rather than contexts of collective or collaborative writing.

When research on genre does turn to studies of young children's acquisition and creation of genre knowledge, they show how echoes of past learning appear in new ways and reinforce the importance of intertextuality[1] as well as dynamism in the study of genre (Wolf & Heath, 1992). In school settings, Chapman (1994) and Kamberelis (1994) show that children do not, in general, learn written school-genres in algorithmic ways. Instead and despite efforts by teachers, children construct texts with

> whatever elements are ready-at-hand in their discursive repertoires....They construct mosaics or hybrids by integrating a variety of text fragments, conventions from a variety of genres, and thematic content that may not typically co-occur in individual texts. Sometimes this process seems quite accidental; sometimes it seems quite intentional....Sometimes children are praised for their creativity...; more often these texts are seen as aberrant, and their child authors are disciplined toward producing more conventional texts. (Kamberelis, 1995, p. 140)

[1] The term "intertextuality" refers to the interweaving of text fragments, ideas, and meanings in the creation of new texts and ideas. Thus, a text is not one text, but made of many. For example, according to Wolf & Heath (1992), "The word 'story' does not define a single narrative confined within the limitations of one combination of setting, character, and plot. rather, it is the intermingling of texts, where relationships exist among characters, similarities abound among settings, and plots that carry associated themes occur" (p. 204). In terms of genre, writers "pick up" textual regularities and apply them in new ways to new texts. These new texts are thus woven from past readings, experiences, and visions of "parallel" or similar texts.

In studies of children and their writing we often see heavy use of intertextuality and creative intent (Daiute, 1993). Daiute termed these appropriations and alterations by young writers for rhetorical purposes "youth genres" in her study of fourth and fifth grade writers. These "youth genres" illustrate to some degree that learning genres is more like learning languages than algorithms (Kamberelis, 1995). Theory and research on learning language continues to confirm that language acquisition is a situated and social activity for the learner (Harste, Woodward, & Burke, 1984; Heath, 1983; Rogoff, 1990). The same can be said for genre acquisition.

Genre studies (e.g., Bazerman, 1988; Chin, 1994; Kamberelis, 1995) have shown quite well that genres are in fact flexible and dynamic. Yet while the descriptions of dynamism are necessary to understanding genres as social actions, *how* these genres shift through social practice has been the subject of less research. The "how" of genre stability and dynamism is located within the social interactions among members of a group--those in charge who "allow" changes; those in the group who "dare" to introduce creativity; those who "toe the line"; those who discipline "aberrant" texts. Who are these members who introduce change? And how do the group dynamics allow for these innovations? On the other hand, who are the gate-keepers of genres and how do they justify their positions? In short, the social practices of learning genres need more attention if we are to be informed teachers of writing.

Knowing how genre learning takes place means shifting our attention for a moment to theories of learning that would enable genre as social action to be learned, acquired, adjusted, and adapted by members of a group or community. When Berkenkotter and Huckin (1993) take a sociocognitive approach to understanding genre, they propose their "activity based theory of genre" by outlining five principles of genre that prove quite useful for framing a theory of learning genre practices. But if we view these five principles in independent ways, we may be back to an incomplete or fragmented understanding of how genres perform their social action as a *system* of

activity. Thus, this study begins to describe the interdependence of Berkenkotter and Huckin's (1993) five principles and the relationships among genres, texts, and social practices (Kamberelis, 1995).

Doing, Saying, and Writing:
An Activity-Based Theory of Genre

As Kamberelis (1995) and Berkenkotter and Huckin (1993) note, genre theory and research has crossed disciplinary lines from rhetoric into a myriad of social science communities. Such blurring of disciplinary concerns and research necessarily lead to complexities in "doing research" on genre. What perspective to take? What aspects to leave out when taking a perspective? What is lost or gained? Numerous theoretical descriptions of genre mention the idea of activity. How this term is defined and described will guide how it might reveal information for the understanding of genre acquisition. Indeed, it is in social constructivism (Vygotsky, 1978), or activity theory, that I have located a theory of learning that supports the visions of genre and critical literacy essential to understanding the social actions created through the student newspaper.

From the socio-historical school of research, anthropologists and psychologists have crossed disciplinary boundaries to explore learning and knowledge as aspects of a socially constructed activity. The school of "social constructivism" established by Soviet psychologists serves as a theoretical foundation for looking at the individual learning within social contexts. This emphasis on activity and social interactions has shifted educators' and others' attention from the individual on his or her own to a participant who is learning in shared goal-directed actions. With respect to goal directed actions, Vygotsky (1978) and the Soviet Activity Theorists (e.g., Leontiev, 1981) posit the eminence of language in the construction of knowledge and learning as individuals participate in collective activities. In reflecting on learning, Vygotsky

(1978) makes vital the connections among thought, language, individual, and community when he writes that a child comes into a social-historical context when "speech becomes rational and thought verbal" (p. 91).

Socio-historical researchers who frame the individual as learner and constructor of knowledge within social situations (Rogoff, 1990; Vygotsky, 1978) posit the idea that knowledge and language are constructed within social activity--goal directed actions that lead to collectively agreed upon material products or ideas. With respect to language genres, individuals learn to navigate in communities by *adopting* the language processes and uses of the community as part of becoming a member of the group. And simultaneously, individuals create independent selves through innovative language acts by *adapting* community language uses for personal goals and social change.

Activity Theory, with its three levels of analysis, allows a close examination of genre, of how language connects the individual to him/herself, connects the individual to a local community, and finally, how an individual connects him/herself to a wider, structured society. It is through language that many of these crucial social connections and interrelations occur (Freire, 1985; Freire & Macedo, 1987). Thus, a study of how language creates these links of self to self, to community, and to society may aid researchers in understanding how language processes can be approached in learning institutions such as schools.

In the past, much of the research on school-based learning has looked past the "insiders" creating meaning out of learning activity (Bakhtin, 1981; Berkenkotter & Huckin, 1993), past students' mutually collaborative construction of social purposes for writing and reading and focused on the individual with his or her text and on the teacher as guide and evaluator (Athanases & Heath, 1995; Corbett & Wilson, 1995). Though social surroundings may be seen to *affect* the writer's choices, writers are mainly viewed as solving individual writing problems through independent writing processes (Applebee, 1986; Doyle, 1986; Flower, 1989; Hairston, 1982; Willinsky,

1991). Still, some researchers on writing share a concern for moving beyond research on the individual engaged with a personal text to consider the complex interrelationships between individuals learning and the social interaction of multiple language acts (Carter, 1990; Fingert, 1994; Flower, 1989; Fulkerson, 1990).

Interest among literacy researchers for understanding the situatedness of writing activity has led to re-visioning of concepts such as literacy and genre. As discussed above, when genre is viewed as the social action taken by a group through the "tool" or means of writing, then understanding the messages of written products is linked to the social activity and meanings of the writers and the writing (Kamberelis, 1995; Miller, 1984). Such a view of genre swings the focus from finished written products on to what the students and writers are actually doing, saying, and writing as they engage in literate activities (Berkenkotter & Huckin, 1993; Dyson, 1989; Harste, Woodward & Burke, 1984; Heath, 1983). Such a focus on activity--doing, saying, and writing--can shed light on the meanings of literacy for participants in literate acts. As genre theorists note, these meanings will bear both the echo of the past as well as signal changes in the present--echoes and changes that reflect the social system (Bakhtin, 1986; Berkenkotter & Huckin, 1993; Fowler, 1982; Kamberelis, 1995; Miller, 1984). Extended to specifically address literacy, then, literacy activities--in this case, genre-making--will be linked to both the individual user and his/her purposes as well as the meaning of literacy that arises out of a wider social world. Thus, from a socio-historical perspective, researchers are able to perceive students as engaging in a "community of learning" (Brown, Ash, Rutherford, Nakagawa, Gordon, & Campione, 1993) where their literacy actions contribute simultaneously to the community's activities--to the collective voice--and to developing their selves as independent, experienced learners.

Situating the Writer within Activity

If an important principle of genre is defined as social practice, the question of learning genres or social actions and practices becomes partly a question of how

knowledge is created: within individual minds or within social groups. Studies that focus on either the individual or social interactions reflect the epistemological conflict that has long divided psychologists from anthropologists or, indeed, psychologists among themselves. Activity theorists have attempted to shift the conceptual terrain to humans engaged in systematically connected, goal directed actions as mediators for human knowledge acquisition (Eisenhart, Finkel, & Marion, 1996). In his introduction to Leontiev's chapter on activity, Wertsch (1981) notes that Leontiev distinguishes between foci of psychologists on the individual as creating knowledge from external conditions or from idealizing from internal mental work:

> Our knowledge of the world is mediated by our interaction with it. On the one hand, this reflects a common criticism made by Soviet psychologists of theories that treat humans as if they were passive receivers of stimuli. On the other hand, it reflects a criticism of idealistic theories that treat humans as if they created their knowledge of the world by conscious reflection. (p. 38)

Leontiev argues for a theory of learning that has at its *center* human activity, especially human activity conducted in collectives where members share motives and goals.

Furthermore, activity theorists criticize the categorical distinctions between the individual and society. Leontiev (1981) defined this problem of the individual and society in his attempts to describe activity as the unit of analysis most appropriate for exploring human learning and understanding.

> We must make a special effort to work against understanding human activity as the relationship that exists between individuals and the society confronting them. We must emphasize this since psychology...stresses the opposition between the individual and society. According to this view, society is just the external world to which the individual must adapt in order to survive....However, this misses the main point that in a society, humans do not simply find external conditions to which they must adapt their activity. Rather these social conditions bear with them the motives and goals of their activity, its means and modes. In a word, society produces the activity of the individuals it forms. Of course, this does not mean that their activity simply personifies the relations of the society and its culture. There are complex transformations and transitions that tie them together. (p. 48)

Thus humans engaged in material activity in collective "object orientation" maintain and "transform" the structures of the activity. In terms of writing the school newspaper, a genre, then, reflects the social conditions under which a paper is produced and "bears with it the motives and goals" echoing from a wider society and generated from "transformations and transitions" made by individual participants.

More recent reflections on the meanings of social constructivism through reviews of its definition and uses mirror again Leontiev's concern that the individual and society are often seen somehow in opposition. In their consideration of constructivist approaches to science learning, Eisenhart, Finkel, and Marion (1996) outline three basic "camps" of constructivist thinking and propose the middle "camp" as being the most providential for researchers interested in the learning and knowledge that are created through social interactions in groups.

According to the authors, constructivism has been defined along a continuum ranging from a stage theory of knowledge construction associated with Piaget to a radical version in which "knowledge is constructed and legitimated whenever it makes sense to an individual in a particular experiential context" (p.277). Between these extremes is a conceptualization of constructivism that is built upon a model of social practice. According to Eisenhart and colleagues (1996),

> In neither of these two [extreme] constructivist frames for addressing science education is serious attention given to the structural characteristics of schooling or science, the social organization of instruction, the tools of language and inquiry..., or the identities that school science inspires. (p. 278)

The features of sociohistorical constructivism listed here align themselves with the features of genre study outlined by Berkenkotter and Huckin (1993). A sociohistorical-constructivist understanding of genre--in this case, of newspaper creation--would put the focus of research squarely on the systematic social practices of the group. Thus an activity-based theory of genre specifies activity as the unit of analysis, as a unit of human interaction that is bounded by shared beliefs and material products. How this

unit of analysis is understood within the framework of genre theory is the subject of the next section.

Activity, Actions, and Tasks: Connecting Principles of Activity with Genre

While efforts to delineate principles of genre such as that accomplished by Berkenkotter and Huckin (1993) can aid in understanding the work or writer, it risks compartmentalization. That is, the five principles as well as the themes identified by Kamberelis (1995) continue to separate the text from the genre as social action or practice. Activity theory provides a way to consider the co-constituent aspects of genre as well as how these aspects interact with real participant-people within a system.

Leontiev (1981) defined Activity systems as arenas of participation in which members of a group successfully negotiate the production of objects and ideas in collective interactions. Activity theory provides an analytic framework that allows for a systematic examination of the interrelationships among genre learning, social interactions, and textual practices. Because understanding the meanings and uses of language in an activity system is a complex endeavor, ways of accounting for and explaining this complexity require "exploring various organizing systems. Different ways of organizing and coding data yield different results" (Athanases & Heath, 1995, p. 268). Thus, a study that uses activity as a guiding concept requires an analytic framework which accounts for the conceptual complexity of the questions asked in the study as well as for anticipated complexity in the data analysis.

In order to accomplish an analysis that captures the dialectic that exists between individuals and their social contexts, several research projects have called upon Activity theory as an organizing construct for their research. Wertsch, Minick, and Arns (1984) drew closely upon the three units of analysis delineated by Leontiev (1981) to show the complexity of literacy learning that individuals engage in as they participate in certain shared reading contexts both at home and at school. For the present study, these three levels of analysis show how the dynamism of genres is linked closely to its

situatedness and social practices; how the form and content of genres reflect and alter the social structures "we constitute and simultaneously reproduce" when engaged in collective activity; how genres signal a community's "norms, epistemology, ideology, and social ontology" (Berkenkotter & Huckin, 1993, p. 478).

Activity: The first level of analysis

According to activity theory, analysis of learning through doing focuses first on the *activity* and takes the broadest view of the interactions among group members. This general understanding of activity situates the activity of participants within the context of wider institutional constraints: "This first level of analysis...provides a bridge to social institutional phenomena" (Leontiev, 1981, p. 55). Activities are linked to motives that reflect a larger socio-cultural context in which they take place. As Leontiev (1981) notes, "in a word, society produces the activity of the individuals it forms" (p. 48). However, he also warns against understanding activity as deterministic, reminding us of the "transformations and transitions" possible in activity as well. This reproducing and transforming relationship between the activity system and the wider social world cycles back to the idea of using language for social action that opened this chapter. Within this context, the students working on the newspaper will simultaneously draw motive for their activity from their understandings of the social world around them and will challenge that larger social world through their genre-making.

In the case of the Zephyr, the activity level of the newspaper production activity system reveals aspects of critical literacy that show how genre-making is linked to wider social worlds. Like activity theorists, critical literacy researchers have also noted the tensions between dominant uses of language that support existing power structures and emergent, challenging uses of language that convey messages of change:

> The notion of emancipatory literacy suggests two dimensions of literacy. On the one hand, students have to become literate about their histories, experiences, and the culture of their immediate environments. On the other hand, they must also appropriate those codes and cultures

> of the dominant spheres so they can transcend their own environments. There is often an enormous tension between these two dimensions of literacy. How can emancipatory literacy effectively deal with this tension so as not to suffocate either dimension? (Giroux, 1988, p. 47)

Through active participation in genre-making literacy activities, members of a group can join with and, or even simultaneously, "emancipate" themselves from larger social constraints. This participation and motivation is best understood within the frame of activity, as "society produces the activity of the individuals" (Leontiev, 1981, p. 48).

Activity as a concept for looking at recurring, goal-directed human interactions captures the tensions inherent in genre: its stability and dynamism, its support and challenge of a wider society's "norms and ideology" (Kamberelis, 1995). Understanding genre as activity, as social action, uncovers these tension-producing elements of critical literacy. Because the activity is a manifestation of larger social interests, students involved with the school newspaper reflect these wider concerns through the ways in which they bring to the newspaper ideas of news, ideas of important stories, ideas of how a paper is to function within the school, and as an artifact of their larger worlds. It is in activity that these larger social concerns are most visible. Thus, within the activity level of analysis are revealed ways in which students echo and reproduce the larger social structures and signal the stability of genres through community ownership of their language forms.

In addition to using the activity level of analysis to consider how the student paper reflects the larger society, the activity can also reveal how students intend to disrupt the larger social traditions and expectations for young people, for newspapers, for social structures. In this way, critical literacy--the giving of voice to less enfranchised people (students)--is linked to the genres produced in the school newspaper activity.

Actions: The second level of analysis

Wertsch, et al. (1984) refer to the second unit of analysis as "goal-directed *action*." This analytic unit shifts the focus from the relationship between activity and

motive within the wider society to actions and goals--the interactions among individual participants that are goal-directed toward a shared product. Leontiev (1981) described the relationship between activity and action: "the basic 'components' of various human activities are the actions that translate them into reality....Any kind of well-developed activity presupposes the attainment of a series of concrete goals" (p. 59-61). In terms of genre, actions can be understood as the social practices (Kamberelis, 1995) that lead to the creation of stable and dynamic language forms that represent the community. When genre is conceived as social action, that is motivated by wider social structures, it will be guided by a "series of concrete goals" that are the results of actions taken by the group.

Berkenkotter & Huckin's (1993) principles of situatedness and duality of structure--that is the constitution and reproduction of genres--require the interactions of participants in goal-directed social practices. Thus the social practices of the group (for example, editorial discussions for topic and authoring selections, editing sessions where the rhetorical value of "correctness" is debated, whole group critiquing sessions, delegation of tasks and roles, processes of becoming fully integrated members, dynamics of power and authority) are situated within an activity system and can be explicitly linked to the goal-directed actions taken by the group.

Critical literacy advocates connect the social practices of the classroom back to the potential acts for critical social action through literate acts:

> Any approach to critical thinking, regardless of how progressive it might be, will vitiate its own possibilities if it operates out of a web of classroom social relations that are authoritatively hierarchical and promote passivity, docility, and silence. Social relations in the classroom that glorify the teacher as expert, the dispenser of knowledge, end up crippling student imagination and creativity. In addition, such approaches teach students more about the legitimacy of passivity, than about the need to examine critically the lives they lead. (Giroux, 1988 cited in McGinley & Tierney, 1989, p. 244)

It may be overstating the case to suggest that students will "examine critically the lives they lead" as a result of participating in literate activity through newspaper writing.

However, it is the case that, when given the chance to render thoughts public through writing particular genres, many students step closer to such examination. A student newspaper run by students and for students is both a social and labor activity within the school that has potential to disrupt the traditional school-based hierarchy by requiring a series of actions in which students examine their own and others' lives in critical ways. Explicating students' "ways with words" as they participate in the recurring social practices of their newspaper can help both researchers and practitioners understand the reproductive as well as transformative possibilities of genre.

Tasks: The third level of analysis

The *"operation or task"* is the third stage of analysis of activity. These constituent parts of actions are embedded in the everyday participation of individuals engaged in the shared activity. This level is more concrete and specific to the work of individuals engaged in meaningful actions toward accomplishment of the collective activity. "How it [the action] can be done...its operational aspect...is defined not by the goal itself but by the objective circumstances under which it is carried out....The performed action is in response to a task. The task is the goal given in certain conditions" (Leontiev, 1981, p. 156). In terms of genre acquisition, the work of the actions necessary for the newspaper production include such tasks as the successful completion of writing processes by individual writers. These processes lead to the "objective circumstances" of the paper production: they lead directly and explicitly to the stories used in the creation of the paper.

Of Berkenkotter and Huckin's (1993) principles of genre, form and content of the social action taken by the news writer would be addressed by this third level of activity. Leontiev pairs operations and tasks with the conditions of the work environment; these are generally perceived as institutional, economic, and personal. Thus, the tasks necessary for successfully completing the goal-directed actions are linked to external conditions, conditions essentially beyond the control of the acting

individual. For example, the technology, classroom conditions, and institutional support all influence how writing processes get accomplished. One condition for producing a newspaper is the dependence on advertising for outside funds. This condition influences, indeed at times dictates, the number of stories and graphics possible in a single issue. Thus one task considered in this analysis is the writers' composing processes as linked to the institutional conditions governing that composing. For this study, the writers' processes are components of the actions of the larger group, as individual pieces are subjected to various types and levels of revision (one-on-one interactions with a peer, group discussions led by a section editor around a piece, teacher-student editing conferences). A writers' writing processes are, therefore, linked to the social practices or actions of the group as a whole.

Research on writing processes is well established though most of it has been done with the individual writer and his/her piece of writing. Linda Flower's (1985, 1989) work considers a person's socially constructed "rules" for the form and content of writing that guide particular writing efforts. Using think-aloud protocols, Flower's research has carefully documented the cognitive processes and internal dialogues that individual writers engage in as they write for particular purposes and has speculated on their origins. This research has done much to shift writing pedagogy beyond the conventional product orientation to include a writer's process--the recursive steps through topic selection, drafting, revising, and editing taken along the way to the final product. Much of the cognitively-grounded writing research has examined how novice writers learn various rules and conventions in order to become experts. Some researchers like Flower, however, note a need to understand the individual who makes writing decisions that seem to "disrupt" socio-historically constituted genre rules. Flower's (1985) work with professional writers and adult composition students has had direct impact on pedagogy as it calls for student writers to have opportunities to write like "real" writers who engage in processes that sometimes do not "fit" with

traditional school constraints. However, most process researchers' understanding of "real writers" generally considers the context of an individual alone with his or her writing and intention (Atwell, 1987; Graves, 1975, 1983; Murray, 1984). Few look at the systematic relationships that are created within a writing community and the impact of these relationships on the writing that is produced.

This process work is closely linked to a movement in writing research that calls for students to be engaged in "authentic" contexts for writing in school. As with Flower, many writing and composition theorists' interest in authentic writing has meant discovering what professional writers do (Graves, 1975, 1983; Murray, 1984). The process-orientation in writing instruction is a direct result of this position (Applebee, 1986; Doyle, 1986; Hairston, 1982). In such an orientation, students learn possible tasks of producing a piece of writing from topic selection, through invention strategies, through drafting, and finishing with editing. A focus on writing process shifted attention from the paper product and enabled teachers to guide students to the product and to assess student performance based on many views of the piece in question rather than only grading the end result. In some ways, the process approach gave teachers a *context* for understanding their students' work. Other instructional results of looking to professional writers as models for student writers include renewed attention to the reasons people write and the audiences for whom the writing is done--a social rhetorical framework (Carter, 1990; Fulkerson, 1990). Moves to creating authentic contexts for student writing derive from research and learning about what writers do outside of school in their real worlds and importing those features into classroom practice.

A limitation of this writing research has been its focus on the individual as a lone writer influenced or affected by the outside world. Very little research has been done on how writers interact in communities and on how those communities limit and extend the repertoires of writers. These questions arise when writing processes actually occur in social settings where the practices of the community members are part

of the writing processes. In his unique work with children in a writers workshop, Lensmire (1994) does document the negative effects of this "authentic" pedagogy, as the social practices and relations among members of the class in fact became sites for peer-based power and authority struggles. Of his year-long study with third graders, Lensmire writes:

> I thought my story would be a different one. I would teach writing for a year in a third grade..., struggling a bit at first to get a writing workshop running...Children would write themselves on the page, move, be heard, in a place that habitually constrained their voices and bodies to teacher questions, to desks. Our workshop would be a little Emersonian democracy; Dewey's embryonic community....Neither workshop approaches, nor the roles they envision for teachers [and students], are so innocent....Writing workshop approaches to the teaching of writing emphasize increased control by students over their own writing activities and texts....[C]hildren's relations with each other became extremely important for their experiences and writing in the workshop. These relations included the rejection, by children, of members of the other sex as partners in collaborative work, and peer hierarchies granting those boys and girls at the top status and influence, and those at the bottom the brunt of teasing and exclusion. (pp. 1-2).

As children were involved in "authentic" opportunities to pursue their own writing topics and processes, the social practices within this community took on paramount importance. According to Lensmire's work, the silencing of some and empowering of other voices greatly affects the stability and dynamism of genres. While researchers have speculated that giving students opportunities to write like real writers would empower them, the social practices of the group actually conspired to limit writing to a few "privileged" genres as well as people.

Lensmire concludes his study by suggesting that collective writing activities might "disarm" the hurtful social practices of the children in his classroom:

> With collective projects, we support certain peer relations, not by intervening at the level of outward behavior, but at the level of curriculum, by directing their attention to a common problem to be solved. Furthermore, collective problems at the center of activity may mean that individual contributions can "add up," contribute towards the knowledge of the group as it tries to solve textual problems, in ways that individual solutions to individual problems do not. There seems a better chance that children themselves will see and acknowledge the contributions of others when those contributions help in common efforts to produce or respond to important texts. (p. 156)

Thus how individual writers go about actually composing and producing their writing is linked to the power and identities of those engaged in the social practices of the community. Lensmire's work is a caution for researchers of writing and genre to attend to not only the texts, but also the social practices of the writing so that the theories of literacy learning don't stretch into the heights with no reality to back them up.

Activity theory provides a multilevel analysis that organizes the idea of genre as social action in such a way as to help discern the intertwined principles of dynamism, situatedness, form and content, duality of structure, and community ownership. The acquisition of genres is inextricably linked to the knowledge, learning, and meaning-making that an activity-group creates from their shared experiences. By using this nested view of participant actions, a clearer understanding of the individual, his/her learning of genre, and the effects of that learning can be linked explicitly to the social context.

Learning Genre:
Participating in Activity as Part of a Writerly Identity

As Lensmire (1994) discovered, the individual language learner, the language user, holds together the three levels of activity. Activity theory provides an account of the learning an individual does as s/he participates in the work of a social group. Taking a socio-historical-cognitive perspective on learning genre suggests that

> learning the genres requisite for becoming a competent speaker and writer within a particular social formation seems to involve much more than the induction into a new set of discursive practices. Rather it entails a process of identity reconstruction--the interpellation within new subject positions and the appropriation of new values, norms, and ideologies. (Kamberelis, 1994, p. 53)

When a group is involved in embedding its values, norms, and ideologies within its genres so as to take social action through those language forms, it is clear that power and authority are involved in these processes. And clearly, within this

conceptualization of genre and learning as social practice lies questions of who "permits" certain expressions, who "disciplines" attempts to press genre boundaries, and who passes genre expectations on to newcomers. Along with questions of who has power, go the questions of how that power is distributed and employed in social practices. It is this distribution of power that influences the stability and dynamic flow of genres as they do their social work. Thus the "membership journey" becomes critical to understanding the social practices that are genre and how those genre practices are part of "identity reconstruction."

When the school setting attempts to bridge the school-world gap by structuring learning within an authentic social activity, as might happen in a student newspaper activity, the connections between genre learning and a literate identity are "situated" within that community. Lave and Wenger (1991) consider identity construction through a "situated learning theory" that focuses on "learning through practice" and implies a process by which "newcomers" become part of a community devoted to common goals. Their term "legitimate peripheral participation" defines the starting point of those joining an ongoing community: "Legitimate peripheral participation refers both to the development of knowledgeably skilled identities in practice and to the *reproduction and transformation* of communities of practice" (emphasis added, p. 55). According to their theory, knowledge and learning connected with a community mark the movement along an identity "trajectory" from newcomers to old timers. In short, the community exerts a sort of centripetal force on newcomers, drawing them closer to the center of community practices and closer to a community identity.

This spiraling of individuals toward convention in speech and written genre practices bears the echo of Bakhtin (1986) who used the term "centripetal" as well to describe the connections between speech genres and social settings. As individuals become practiced at language functions and forms, they are drawn toward a "center" of understanding the shared uses for language. These patterns of discourse are speech

genres and mark an individual's "membership" in a discourse and social community. According to Lave and Wenger (1991), the knowledge growth and subsequent creation of identity is manifest in "relations among practitioners, their practice, the artifacts of that practice, and the social organization and political economy of communities of practice" (p. 122). Learning and knowledge of genres are thus "visible" in the identities, texts, and social practices of a group or community.

Recently, Lave and Wenger's (1991) apprenticeship model has gained currency in empirical research--especially with respect to out-of-school contexts. Certainly, in the cases of tailors and navigators cited by Lave and Wenger to concretize their theory, novices are incrementally brought into the social community of mature practitioners. Though they allow for the creative potential, cultural "transformation," of newcomers as peripheral members of the group, their main focus is upon novices acquiring the skills and processes necessary to craft products like old-timers. Much can be learned about the individual acquisition of knowledge while participating in peripheral tasks that take an individual toward the "core" of a group, but more can be said about how the newcomers help shape, change, and de- and re-construct the group's activities and products.

In a school setting, few students act in truly peripheral ways, that is, in apprenticeship "ways" to teachers. Institutional demands often require each to demonstrate his or her own knowledge without the benefit of *apprenticing* toward mastery of a skill or strategy. However, in other types of school-based courses, for example the student newspaper, the curriculum has room to allow for the gradual taking on of greater and greater responsibility for production of the group's goals, and the learning trajectory possesses certain features that are similar to an apprenticeship-type model. In the view of such a model, the nature of this growth toward the genres of the group will be intimately linked to the social interactions of the members of the group. As students develop identities as members of the news staff, they adopt and adapt, they

both "reproduce and transform," knowledge of what it means to write and be a writer, they take on identities with power and authority.

Indeed, identity formation is much more than the move toward a "center" of group work, it has much to do with the "stories of self" (Eisenhart, 1995), power structures and social interactional practices that "permit" such membership and identity construction. Researchers of identity formation through active participation in socially defined activities, such as work, have considered these membership journeys. In her long-term ethnographic work with a nonprofit conservation corporation (CC), Eisenhart (1995) considered the "stories of self" that two newcomers told to describe their reasons and growing involvement with the group. While Eisenhart's interest was in how individuals change culture, rather than on how culture shapes an individual, there are theoretical links to genre theory and acquisition and how that acquisition is part of reconstructing one's identity as a writer for the newspaper.

For Eisenhart's (1995) informants, taking on a CC identity was linked to the ways in which he and she "took on" the roles, work, relationships, and ways of interacting they were given as well as those they created at CC. She concludes by noting that "the stories they tell are mediational devices that enable certain kinds of newcomer experiences and disable others; they affect how the newcomers are treated by others, and they anticipate the kind of identities available to them in the organization" (p. 20). These "stories of self" are also a tool to understanding the individual's perspective on social practices and the ways of access or denial to access to the community.

On a school newspaper, adopting and adapting a writerly identity will be part of the participation structures of the group and will be reflected in and will reflect back the genres used in that community. And if "learning is conceived as changing forms of participation in context, or changes that transform newcomers into old-timers, outsiders to insiders, or amateurs into experts" (Eisenhart, 1995, p. 22), then learning genre is in

part to be transformed from newcomer to old-timer, to reconstruct an identity that is reflected in taking on and changing forms of language-use appropriate for the community.

Conclusion

An activity-based theory of genre, grounded in activity theory as defined by Leontiev (1981), allows for the complex of principles guiding genre and genre knowledge acquisition to be considered within a system of text, social practices, and embeddedness in the wider society. These three levels of activity analysis permit a nested view of genre: the activity as linked to wider social structures and values, the actions linked to the social interactions and practices of the members of the community, the operations linked to the processes of writers as they put social action into words.

Social action through language is a value shared by many writers, researchers, educators and students. The words from Toni Morrison (1993) which opened this chapter still ring true: knowing law-language, military-language, hurtful-language as well as healing- and empowering-language means knowing genres as social action. Empowering students to speak and write means showing them the power of genre through social practices and consequential speaking and writing.

Through observation of and conversation with students who found a way to be heard within their school community by writing for their school newspaper, I found a context for the development of writerly identities and a site for critical literacy practices. This group meets and interacts daily in both formal and informal social practices that help define and challenge the messages they send out into the world. Through their words both spoken and written, I came to understand their literate worlds and learned how language "protects" and enriches the lives of these "insiders." For, it is through naming the world that students learn to *"hold the name in mind,...reflect on its meaning, and imagine a changed world* (Berthoff, 1987, p. 1).

CHAPTER 3

DESCRIPTION OF METHODS

"The billiard ball does not make sense of its environment. But the human actor in society does, and different humans make sense differently" (Erickson, 1986, p. 127).

In his seminal chapter, "Qualitative methods in research on teaching," Erickson (1986) takes a literary tack in discussing the necessity for research on the sense-making of students and teachers in classroom environments. In addition to the playful metaphor of billiard balls, he invokes a literary history in the growing interest in common people from Beaumarchais' Barber of Seville to Grimm's collection of German folk and fairy tales as warrant for hearing the voices of the less-heard in educational research--the teachers and students in classrooms. For Erickson, as well as other literary and social researchers, naming the world in writing through characters, plots, settings, and themes--the stuff of data as well as literature--allows the reader to hold a familiar yet strange world "in mind [and] reflect on its meaning" (Bertoff, 1987, p. 1) It is thus the job of the researcher to faithfully report and interpret a society, a community, so as to represent the different sense-making humans, not the billiard ball, in action and interaction with others.

Research on teaching and learning within classroom contexts, as Erickson (1986) notes, is as messy with contradictions and paradoxes as it is seamless with systematic rules and uniform behaviors. These conflicting interactions make up the whole of a community and provide challenges to researchers to discover "what is happening here, specifically? What do these happenings mean to the people engaged in them?" (p. 121). In the case of the Zephyr, my questions began broadly with "what is

happening within this community of writers? And what does this mean to the members?" While these general questions are critical to any researcher at the outset of a study, a more refined set of questions quickly emerges from the junction of participant observation and researcher conceptual interests. In this way, my study is not radically intuitive or inductive and aligns with Erickson's view of "fieldwork as a process of deliberate inquiry in a setting" (p. 121):

> Fieldwork methods are sometimes thought to be radically inductive, but that is a misleading characterization. It is true that specific categories for observation are not determined in advance of entering the field setting as a participant observer. It is also true that the researcher always identifies conceptual issues of research before entering the field setting. In fieldwork, induction and deduction are in constant dialogue. (p. 121)

The idea of constant dialogue between induction and deduction can be extended to the continual conversation between a researcher's interests and the actions and activities of the field setting. Because of pilot work done for a graduate seminar in ethnographic methods (LeCompte, 1994) and because of my own tendency toward literary interpretation of "real world" actions, I entered the field setting of the Zephyr with the "conceptual issues of research" that I outlined in the previous chapter. As *I* tried to make sense of the sense-making the students engaged in through social interactions around their texts and community, I entered a world of belief and doubt (Elbow, 1986) that led me to move back and forth between taking-in and believing in the scenes and stepping back with a more analytic eye. This back and forth allowed me to be both "*naive realist*--believer in the taken-for-granted reality [I] perceived at first glance" (Erickson, 1986, p. 126) and systematic analyst piecing together the meaning of the scene from the perspectives of the participants as evident in their daily classroom interactions, their writing, and their relationships with a larger society.

Because the research questions guiding this study are specifically linked to social practices and community artifacts, a search for the meaning-perspectives of the participants requires a naturalistic and qualitative approach toward answers. Thus, I have drawn my research methodology from qualitative analytic approaches used by

Activity Theorists such as Leontiev (1981) and Wertsch (1981); social scientists like Dyson (1989; 1993; 1995; 1997), Erickson (1986), Heath (1983), Rogoff (1990), and Tannen (1984; 1989); and rhetorical theorists like Bazerman (1988), Berkenkotter and Huckin (1993), Elbow (1986), Fulkerson (1990), and Kamberelis (1995). The methods employed by these researchers have given me a framework and concrete ways to answer questions about the links among social practices, texts, and genre in the classroom setting of the Zephyr. The ways in which I invoked specific analytic tools are detailed in the data analysis section in this chapter.

In the following sections, I set the stage for my study by describing the site, participants, and newspaper product; I define and describe the kinds of data I gathered; and I outline my steps of analysis illustrating with brief examples from the multi-perspective, triangulated data that I collected (Sevigny, 1983). This chapter concludes with a preview of the next three chapters of findings.

Where? Who? and What?

Description of Site, Participants, and the Newspaper

Locating a site for this study involved negotiations between my biases about literacy learning and teaching and the realities of public school education. My definition and understanding of literacy as a tool for social and personal agency led me to explore contexts in which literacy was being used in ways that were non-traditional and socially connected. This understanding of literacy is borrowed from many current literacy researchers, and is part of Willinsky's (1990) meaning of "the new literacy." My bias in defining literacy as a social act follows from the work of Heath (1983), Rogoff (1990), and Willinsky (1990) and has been well-articulated by Dyson (1997) in her book, Writing Superheroes: Contemporary Childhood, Popular Culture, and Classroom Literacy. About her work she says:

> I hope to contribute to and extend sociocultural visions, which portray learning to write as learning to use the medium to participate in cultural

> life in socially appropriate ways. I illustrate that children's ways of writing are shaped, not only by their interaction in adult-guided worlds, but also by their social goals and ideological positioning in peer-governed ones. Moreover, social identification and social conflicts, not only in social interactions, make salient new kinds of writing choices, newly imagined ways of depicting human relationships. (p. 6)

Thus, with Dyson (1989; 1993; 1995; 1997), Heath (1983; 1991), Rogoff (1990) as company and support for my own understandings, writing and literacy are both words on a page as well as actions in a social world. Like Dyson, I hope to "extend sociocultural visions" of literacy learning. Finding a site in which to explore the links between the skills of writing and the social actions of written genres was paramount for this study.

I set out, then, to find a classroom in which writing was used for real social purposes, where students had ownership and decision-making opportunities, where the teacher played a facilitative as well as instructive role, and where the everyday activities of the classroom concentrated the efforts of participants on a shared goal simultaneous with writing production. A news-classroom could ostensibly be such a location. Thus, I knew the kinds of literacy practices I was looking for--in theory--before entering the classroom for research purposes.

Through a series of contacts made to both the local and outlying school districts, including teachers, vice principals, and assistant superintendents, I proceeded with a process of exemplary sampling (Heath & McLaughlin, 1994; LeCompte & Preissle, 1993) to find a strong site for this study, one that enacted my definition of literacy practice and where the teacher welcomed research in the classroom. I engaged in three types of sampling: telephone conversations based upon word-of-mouth reputation (four sites seemed possible), selected classroom observations (three sites were amenable; in the fourth the teacher discouraged research in her classroom for academic year 1995-1996), and reading the respective newspapers. Three of the four possible sites had newspapers that published at least once a month; less frequent

publishing would have limited the number of activity cycles to which I could have been privy.

As I built my phone tree through contacts and began my exploratory observations, one name kept appearing--Mr. Tillinger of Langston High. Mr. Tillinger is the teacher facilitator for the Zephyr, the student-run newspaper created and produced at Langston. The Zephyr has the distinction of being a course-for-credit activity, has a long history of financial self-sufficiency through advertising sales, and had won the 1995 best student newspaper award from the Colorado High School Press Association.[1] When I approached Mr. Tillinger with the possibility of him and his students participating in my thesis research, he was most interested and enthusiastically invited me to observe. I went twice before selecting this site as my choice. A handshake and exchange of permission letters and human research subject consent letters solidified the agreement.

Where? A Description of the Site

The site for my study, Langston High School, is located at the center of the mid-sized town of Langston. The town population is approximately 60,000 people and is located on the Front Range of the Rocky Mountains. It is a mix of plain, rural life and shiny, high technology industry.

Langston High is part of a district which generally practices a progressive education and offers a variety of educational options to parents through specialty programs, schools within schools, and alternative formats. At Langston, subjects are taught in "blocked" times of 90 minutes a day. Graduation of approximately 400 students each year is earned through credit accumulated in courses designated by the district curriculum board. About 1500 students drawn from the town and a wide-ranging rural area attend grades nine through 12.

---

[1] Continuing with their reputation for quality newswriting, the Zephyr also won this award the year following my research, 1997.

The newspaper itself is produced on Macintosh computers in Langston's Writing Computer Lab. Layout of the paper is done on paste-up day first on Quark Express, then by hand in room 122. The final days of production are hectic and work often spills out into the cafeteria or other rooms where pizza and napkins litter the desks and floors. The Zephyr is printed by an outside printing source (Rocky Mountain Printing), and copies are bundled and distributed school-wide first thing on Friday mornings.

Who? A Description of the Students and Teacher

The 23 students who worked on the staff of the Zephyr during the time of my study ranged in age from 16 to 18 years old. They were a combination of girls and boys, of AP-class-takers and slackers[2], of mainstream-public-school-goers and alternative-home-schoolers. Jodi, a home-schooling junior, explained that the newspaper class was the only one he took at the "real" school.

> I hated eighth grade. I hated school. I was going to quit altogether. Then I did the home school thing--actually it's called a program of Walk About, like the Aboriginals in Australia. I came here because sometimes there are classes that I *want* to take. Like this class, where I get to do my photography and writing for a real reason. But that's how my whole education is now anyway, I get to do my work for real reasons that are important to me. (2/2/96)

The people I came into contact with on the Zephyr represented a wide range of school students and largely defied broad generalizations about the make up of the class. In the following paragraphs, I present a profile of this community from a number of different perspectives.

Table 3.1 provides a profile of the class and Zephyr staff. I selected the descriptive category headings as a way to begin exploring the dimensions of difference

[2] I learned about the word "slackers" and other peer group designations throughout my study. My sources for the list in Table 3.1 include the students in the class, students not affiliated with the newspaper, and a Zephyr article discussing cliques in the school (Zephyr, 11/17/95, p. 13).

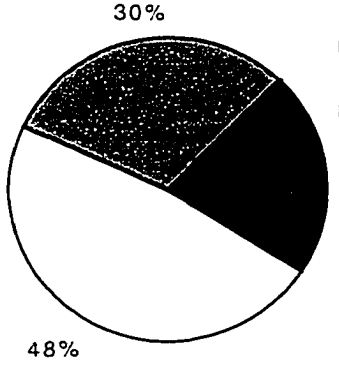
Table 3.1

Profile of Participants on the Zephyr Staff

Name	Gender	Yr	Time on Staff/ Returning	Position on Zephyr	Peer Membership	Future	Rank*
Amie	F	10	1 year yes	Reporter	Alternative	Photo-journalist/ Free lance	M
Andi	F	11	.5 year no	Reporter	Straight	College then Graduate School	M
Annie	F	11	2 years yes	Reporter	Popular	Don't know	L
Cass	F	10	1 year yes	Reporter	Skater	Free lance writer	L
Cris	M	11	1.5 years yes	Reporter	Straight	College	L
Dano	M	11	1 year yes	Reporter	Straight	College	L
Dave	M	12	2 years graduate	Reporter; Feature Editor	Straight	College for pre-medicine	H
Dean	M	11	2 years yes	Sports Editor	Popular; Jock	College	M
Jenne	F	11	1 year yes	Reporter; Cartoonist	Skater	College	M
Jesse	F	11	2 years yes	Reporter; Editorial Editor	Alternative	College	M
Jodi	M	11	1 year yes	Photographer	Alternative	Don't know	L
John	M	12	4 years graduate	Reporter; News Editor	Jock	College	H
Josh	M	11	2 years yes	Graphic Designer	Alternative	College for film and music	M
Julie	F	12	1 year graduate	Reporter	Popular	College then Graduate school	H
Lori	F	12	1 year graduate	Reporter; Circulation Manager	Artist/ Dancer; Popular	College for dance	M
Louis	M	11	1 year yes	Reporter; Entertainment Editor	Death Rocker	Don't know	M
Mark	M	12	5 years graduate	Reporter	Conservative; Straight	College	H
Mel	F	10	1 year yes	Reporter; Page Composition	Straight	College	M
Nic	F	12	1 year graduate	Reporter; Advertising Manager	Cowboy	Music performance	M
Rusty	M	10	1 year yes	Production Mananger; Computer consultant	Popular clique; Youngster	College	M
Sera	F	12	2 years graduate	Reporter	Popular; Jock	College	H
Shawn	M	12	1.5 years graduate	Reporter; Business Manager	Straight	College	M
Stuart	M	10	1 year yes	Reporter	Jock	College	H
Troy	M	12	3 years graduate	Reporter; Editor-in-chief	Popular	College then Graduate School	H

* Ability ranks are the consensus of student and teacher determination. I asked students to self rank and asked the teacher's thoughts on the ranking.

and similarity among members of the staff. This brief profile of the class shows the mixture of ability and peer group membership. The charts and figures that follow summarize some ways to view the various groupings among staff members.



Proportion of students of low, middle, and high range ability. Ability labels were negotiated between student and teacher with a very high level of agreement. Only three of the 23 designations had to be discussed before the student and teacher agreed.

Figure 3.1. Proportion of high, middle, and low ability students

Nearly half the class identified itself, with very high agreement from the teacher, as middle range ability students. The other half was almost equally divided between high and low ability students. Those who described themselves as low noted their disconnectedness with school and a desire to just "get through it."

The ability labels are only one way to see variety in this classroom. Another way is more closely linked to the students' own community--clique labels. The idea of cliques or exclusive peer groups is not foreign to those who do research on adolescents (Varenne, 1983), and the community in which the Zephyr exists is no different. In the November 17, 1995 issue of the Zephyr, reporter Cass interviewed students around school for their views on cliques and offered the research and professional take on such groups. She quoted students who explained both the value and limitations of peer group cliques. This was similar to the descriptions and explanations about peer groups that I was hearing in the Zephyr classroom. For instance, the labels for cliques used by the students in the article were the same as those those used by Zephyr members. Cass

quoted a sophomore as saying, "In school, there's jocks, gangsters, cowboys, alternatives, skaters, straights....But when you're not in one, you can learn a lot more about different people" (p. 13). The terms "jock, cowboy, alternatives, skater, straights" were all used--in addition to others--by Zephyr staff members in reference to their own and other's peer group membership. Figure 3.2 shows the self-designations of staff members.

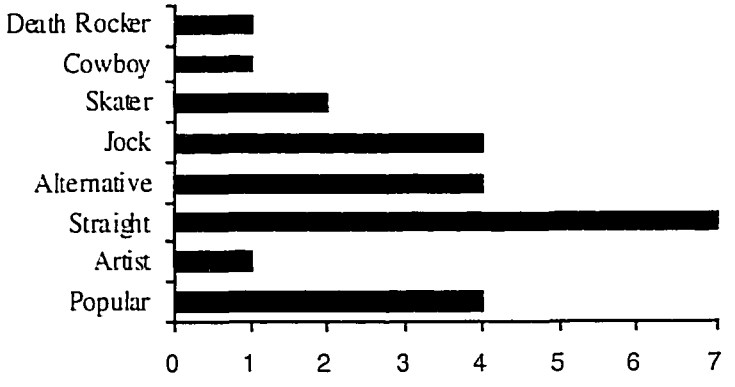


Figure 3.2. Peer-group memberships for Zephyr staff.

These peer group designations illustrate the wide variety of types of students on staff. The terms, "popular, artist, straight, alternative, jock, skater, cowboy, death rocker," are all taken from student language. I came to understand these terms in the course of interviewing, informal conversations, and observing students as they interacted with one another during the class sessions.

Second, Cass' article noted that categorizing people is easier in the abstract than it is in real life. She quoted a junior as saying, "There are a lot of people I couldn't find a specific group for, and some would fit into a lot" (p. 13). Cass' article goes on to note the importance of membership in some group to avoid loneliness and isolation, but notes that "Most students today do not want to consider themselves part of a clique" (p. 13). One theme that was central to the ways students described their experiences on the Zephyr involved the successful mixing of peer groups within the newspaper production

class. In keeping with Cass' message in her article, "Birds of a Feather" (Zephyr, 11/17/95, p. 13), though all Zephyr staff self-identified with a group, 82% insisted on caveats to their memberships in these groups by noting that they might be seen as belonging in more than one group and that, ultimately, they really didn't like to categorize themselves.

Along with variety in ability and peer group membership, two other dimensions of difference show the "mixed up" aspect of the class: grade in school and gender. Figure 3.3 illustrates the distribution of boys and girls across grades.

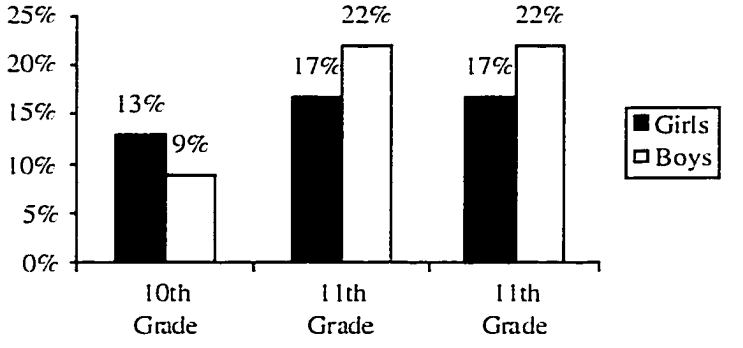


Figure 3.3. Proportions of girls and boys according to grade

The boys and girls were almost equally divided with 12 boys and 11 girls. The distribution across grades held fairly stable, with a slight edge to the boys as veterans of the group and a slight edge to the girls as newcomers. Because most members joined as juniors, 11th grade, the lower number of 10th graders was not unusual.

One final aspect of diversity found in the classroom is the time on staff. Typically students joined the staff as juniors, though some years, as Mr. Tillinger pointed out, "are heavy in sophomores or seniors. When it's seniors, you always worry about the next year's staff and it's youngness. When it's sophomores, you feel sorry for the seniors who have the age advantage, but not always the experience advantage" (5/14/96). The average stay on the staff was just under two years. This

longevity of membership helped assure a relatively stable passing of custom, tradition, expectation, and expertise from one staff to another. It also helped stabilize the various peer groups by allowing for enough time for staff to get to know one another and become comfortable with the daily practices of the group.

The point of presenting these varied dimensions of similarity and difference among students is to corroborate staff impressions that "there are a lot of groups in this class, but somehow we manage to get along" (Cass, 3/18/96). Diversity of peer groups in the class was of intense interest to the students in the class, and contributed to the development of the "teamwork" and "family" theme that ran throughout the data. Even the Dean of Students noted that,

> The newspaper seems different from, say, student senate, where the kids are all the same, or at least the same ones appear from one year to the next. And they are usually from the popular crowd. On newspaper it seems like a larger variety of types of students are attracted, and not always only the successful ones. (5/21/96)

Thus the busy atmosphere of the classroom was constructed from the woven interactions of students from various ages, grades, peer groups, and levels of academic ability. This "mixed up" class contrasts with traditional high school classrooms where groups tend to be much more homogeneous (Applebee, 1984).

Mr. Tillinger: Teacher and Managing Editor

The participant profile would not be complete without attending to a central figure--Mr. Tillinger, the faculty sponsor and teacher of the class. In the following paragraphs, I offer a fairly comprehensive profile of Mr. Tillinger by drawing on interview and fieldnote data. I have made this unlikely choice to include descriptive data on the teacher as part of my methods chapter because most of chapters 4, 5, and 6 focuses on the *students'* interactions and writing rather than Mr. Tillinger's teaching. It is important to note, however, that the class functions as it does under the careful direction and planning of the teacher.

Having worked with young people for 18 years, Mr. Tillinger speaks easily with any student who comes into his classes and has a reputation for being tough but fair. As Sperry, former editor of the Zephyr told me, "He's the kind of teacher you might have hated in school, but look back and realize how much you learned. But most people didn't hate him. He commanded respect" (2/13/96). Mr. Tillinger had many opportunities to evoke these feelings in students as his teaching load included two ninth-grade English classes, an Advanced Placement History class, and the journalism class that produces the Zephyr. This schedule kept him quite busy planning, instructing, and grading along with developing relationships with students.

For Mr. Tillinger the key to a successful class was rapport which meant making connections. In fact, a constant in his language about teaching was the word "connection." On one informal occasion, we sat together and talked about why we like to teach and what we accomplish for ourselves:

> What's cool in my mind is the students aren't working for the A. The same with this job. I'm not working for the pay. I get paid the same whether I do the job or I don't. So if I can look myself in the mirror and say I'm really doing the job, then that means all the world. That's about making connections. And my whole life flourishes. (5/23/96)

Mr. Tillinger talked of making connections with individual students, of observing connections made among students, and of pressing youth to make connections with the outside world. "When they do interviews and talk to people they don't know and explore the world that way, they get this new sense of self. It's a professionalism that is key" (5/23/96). Connecting with self, others, and a wider world echoed through Mr. Tillinger's talk about his own participation as well as about the participation of the students. These connections formed the emotional glue that held the diverse students and Mr. Tillinger together.

In reflecting on his 18 years as a Managing Editor, faculty sponsor, and teacher for the Zephyr, Mr. Tillinger noted the continued renewal the paper provided for his teaching and connection with students. Mr. Tillinger, himself, was a columnist for the

local professional paper, The Langston Tribune. And he compared his emotional attachment to his own writing with his feelings for the Zephyr:

> I have an ego attachment to that column. Who wouldn't? So for the first six months of that column, I probably cared a lot about how that column was appearing, anxious to open up the paper and see that it had made it....I still put in the effort in writing the column now, but as far as whether I ever even look at it when it comes out is nowhere even as exciting as when the Zephyr comes out. When the Zephyr comes out, and this is amazing because this is 18 years of doing the Zephyr, and I should be burned out and I should be tired of it and so forth. But I get excited every single issue because there's something in there we haven't ever done before, or it's a different look, or it's a kid's first big piece. The kids are excited. I'm excited. (5/23/96)

Just as quickly as he spoke of the excitement of the paper, Mr. Tillinger spoke equally seriously about his emotional commitment to his students--the source for his excitement about the paper. "I wouldn't care about the paper if I didn't care deeply about the kids" (5/23/96). Examples of his caring abounded in the classroom through small group interactions, one-on-one conversations, and through both formal and informal episodes of teaching.

On one occasion I observed Sera, a veteran writer for the Zephyr, confronting Mr. Tillinger with her perception of being treated as childish. As Mr. Tillinger listened, Sera's voice rose, turned to tears, and finally ended in a door slamming. He spoke quietly with Troy, the editor, and excused himself from the classroom. I watched as he followed Sera down the hall and engaged her in conversation. While all was not perfectly ironed-out, they did achieve some reconciliation. He connected this incident with his general beliefs about working with the students on the Zephyr staff:

> I get so wrapped up in these kids. I love it when we're on task, and we're flowing smoothly, and we get a chance to talk politics, religion, or I get to know these kids better. We're gossiping....I love these kind of insider talks. I feel like I'm on a friendship level with these kids. I'm connected with a whole bunch of former students who were kids. And I'm not married, so that's kind of my support....I don't get an opportunity to be a parent, and I don't want to be the parent of any of these kids. But I do get to be kind of an older brother, older adult advisor, that sort of thing. If I can help some kid with some specific problems and it matters a lot to me. It's easy to connect. If there's a falling out, if someone's mad at me, it's hard to take. It's hard to take. You witnessed one this year with Sera....We have such a long

> relationship going back to last year....It wasn't easy at all to work that out....But that's more the aberration. That's not the usual connection I get. (5/23/96)

Students corroborated Mr. Tillinger's perception of an emotional tie between kids and teacher. Often they spoke of his honesty and willingness to discuss difficult issues. Troy noted that "he can be one of us, or he can be a teacher. It's a great combination, one you don't get much of in high school" (4/12/96). As in any family, Mr. Tillinger took on multiple roles--disciplinarian, motivator, confidante, parent, sibling, friend--depending upon the present interpersonal context.

His ability to shift with relative ease among roles and emotions was born of experience and profound belief in news writing as a medium for social action. Indeed, Mr. Tillinger's personal history spoke volumes about the sort of experiences that contributed to his flexibility in the classroom. He identified himself as the end of a long family line of newspaper writers--all affiliated with Langston High; all affiliated with the Zephyr:

> It wasn't called that [the Zephyr] back then, but my grandfather, my father, and my aunt were all on the paper at one time or another.[3] In fact, my aunt was the business manager the year the paper folded. Me and a buddy were the ones who brought it back after that. (1/30/96)

His reasons for restarting the paper were political. As a senior, after asking an English teacher to be the faculty sponsor, "a guy we knew who would leave us alone and stay in the teachers' lounge" (1/30/96), he and a friend set about reasserting the presence of the Zephyr in the school by engaging in some "in-your-face reporting" (1/30/96). Indeed, this iteration of Langston High's newspaper ended after only one semester of circulation because of political pressure:

> It was just the two of us. And it was in the 60's and 70's when things were so political. The paper was a really political thing for us. But the

[3] Mr. Tillinger's family history with the newspaper began with his grandfather's membership in the 1930s and continued with his own father's in the late forties. As he noted, his aunt's membership, 1962-63, marked the end of an era when financial difficulties forced the Zephyr to shut down. In 1968, with support from the administration and the willingness of an English teacher to advise, Mr. Tillinger and a friend resurrected the newspaper.

> advisor didn't care. He didn't pay any attention to us until we put a picture of a starving Biafran on the front page with a poem in an issue right after Thanksgiving. After the advisor talked to the principal, they told us we could take an F in the course, or quit. We decided to quit because we had also started a teen page in the Langston Tribune. (1/30/96)

His high school experiences on the failed Zephyr did not deter him from an undergraduate degree in journalism from Northwestern University, a variety of stints writing for several newspapers, and finally the decision to become a teacher "so I can show them how important the newspaper can be to their lives and to have some time to write a novel" (1/30/96).[4] Mr. Tillinger brought to his teaching his deeply held belief in the press as a political tool, as an agent for change in a democratic society. He also brought an understanding of the dynamic nature of social, political, and interpersonal relationships.

Mr. Tillinger also described his commitment to personal growth and life-long learning as crucial for avoiding burn-out as a veteran teacher:

> I'm always learning about my own relationships and how to teach. I one time was on a hike with a banker, a friend of my brothers. And we were on this hike way above timberline, talking about life, talking about jobs, and he asked me how long I had been teaching....and he says, "You must be starting to get it down." ...Well, it's weird, because every year I'm starting the school year and I think, this is the year I'll have it wired. And just as I'm thinking I'm on the brink of having it down, and somethings work; somethings don't work. And I'm always learning and growing. So, I'm always having to do self-examinations....With that episode with Sera, what can I learn from that? I grow from that experience....It's all about making those connections and thinking about how to grow and learn. (5/23/96)

By focusing on his own learning, Mr. Tillinger brought to the class an approach to teaching that involved the whole class in teaching each other and an open-forum for ideas. His frequent emphasis on *talking*, about life and jobs, was a prominent theme in his language about both personal and professional teaching and learning.

---

[4] As of July 1996, Penguin Press had agreed to publish a mystery novel that Mr. Tillinger and his sister co-authored. The yet-to-be-titled novel is due in bookstores sometime in 1998.

In Mr. Tillinger's classroom, language was the recognized way to communicate, for communication created connection.

> We have to talk. What's worse is not to talk. Whatever it is. And if I can get across these little things and not put them on the defensive or make them nervous so they can't respond, then we can talk and work it out. It's the same for them. They have to figure out how to talk to each other and work it out too. (5/23/96)

Mr. Tillinger's focus on language and talk echoes others' research on the social construction of knowledge through linguistic interaction (Heath, 1983; Rogoff, 1990). It also reflects research on the centrality of talk to personal relationships and public communities (Dyson, 1989; Tannen, 1984; 1989). Thus, Mr. Tillinger's classroom was an arena for talking and writing--illustrating both a personal and professional commitment to cultivating language for learning.

In all, Mr. Tillinger brought to his work a history of writing for newspapers for social purposes, a future of writing goals including novels and continued work for the local press, and a powerful belief in the importance of the spoken word to create connections and relationships among all members of the classroom community. These experiences with writing, speaking, and teaching created a stable foundation of credibility for students as they worked for the Zephyr.

Finally, in addition to the students and Mr. Tillinger, my research participants included the principal, Dean of Students, two other English teachers, and nine other students reading the newspaper but not affiliated with its production. These people helped me establish the newspaper as an artifact of this school community. These more secondary sources of data also helped me see mutual influences on the Zephyr: the influence of these people on the paper; and the influences the paper had on the Zephyr's reading audience.

What? Description of the Newspaper, the Zephyr

The Zephyr newspaper with a circulation of approximately 1500 (including an out of school subscription list of 430 that included alumni, parents, and advertising

patrons) arrived off the presses once every three weeks. Staff members distributed the paper to high traffic areas on the Friday morning of publication. These areas included: the cafeteria, the gymnasium bleachers, the main office, a copy in each teacher's and administrator's mailbox, classrooms, and three, non-coin-operating newspaper boxes located at hallway junction. A readership survey conducted by staffers Jesse and Paul (Zephyr, 5/3/96, p. 2) found the following readership statistics:

- 65% of the student body reads all or some of the paper each time it comes out
- 87% of the student body has read portions of at least a two issues over the year
- 45% of the student body reads all of the paper, all of the time

The figures from their survey reassured the students that they were distributing the paper in places where students could and would get it. The figures also assured them that there was a real reading public for their work.

The Zephyr uses a tabloid format and typically runs from 18 to 24 pages (The Feb. 23, 1996 issue of 28 pages set a Zephyr record for length). The contents of the paper are organized into five sections: News, Features, Editorials, Sports, and Entertainment. These section types have been stable for the past 15 years. Each section has an editor who assigns, collects, edits, and arranges the stories, photographs, and graphics in his/her section. Each editor works with the staff illustrator, staff graphic designer, and staff photographer to assign and incorporate visuals into his/her section. Section editors also work with Mr. Tillinger and Troy, the Zephyr editor-in-chief, to assure a balance of topics within each section. Of greatest concern is a balance between in-school and out-of-school topics. Thus, the feature section might include four or five features written on school or teenager issues and several focused on local, state, or national issues. The same holds true of news, sports, editorials, and entertainment.

Reporters with at least a year's experience typically wrote at least three stories for each issue. Newcomers wrote one or two stories for their first two issues--at least

one of which was co-written with a veteran member. Writers were not limited in the number of stories they were allowed to write; however, space in the paper depended upon the advertising budget. While most operating costs were covered by yearlong contracts with regular advertising customers (Planned Parenthood and a local community college were such regulars), staff members still had to seek additional dollars from other advertisers. Issues where staff were unable to procur this additional support were shorter; as in real world publishing, fewer dollars meant less fewer pages. Mr. Tillinger expected all members, including newcomers, to participate in the advertising sales. He noted that, "it makes them all aware of the realities of running a business, and it gives some a taste for that kind of work. And it gives them practice at negotiating" (5/23/96).

The Zephyr had a fairly stable look. Some descriptive characteristics are given in table 3.2.

Table 3.2

Descriptive Characteristics of the Zephyr*

Characteristic	Average for characteristic	Proportion of section stories	Location in the paper
Number of News stories	9 stories	22%	pages 1 to 3
Number of Feature stories	14 stories	34%	throughout
Number of Editorial pieces	5 pieces	12%	page 4
Number of Entertainment stories	6 stories	15%	2 to 3 pages after middle-fold
Number of Sports stories	7 stories	17%	last 2 to 3 pages
Total number of stories in each issue	41 stories	100%	

*Numerical data were determined by counting and averaging from five issues of the Zephyr.

Stories in all sections ranged in length from 6 column inches (see figure 3.4), for a very short story, to 40 inches of text on a single page that included two small pictures (4 in. X 4 in.), to approximately 80 inches for a "double truck"--a large, single theme series of stories in the middle-fold (the "open-book" inside sheet) of the paper. On average, a typical story ran between 15 and 20 inches.

The paper was arranged in a predictable sequence beginning with the front page. The front page always included a left-hand column-wide table of contents that announced the location of selected stories, two news stories continued inside the paper, a photograph on or above the fold, and, of course, the banner heading with name, volume information, address, and date. Selection decisions for the front page fell to the editor-in-chief in consultation with Mr. Tillinger. News stories appeared on pages one to three; page two was always school news. Editorials appeared after news stories on page four and sometimes took up two or three pages. Entertainment always took the two or three pages after the center-fold double truck. And Sports fell to the last 2 or 3 pages and always included the back page.

The predictable sequencing allowed readers to quickly find their favorite types of stories. The readership survey conducted by Zephyr news staff confirmed reader preferences and the need for relatively stable section locations. According to the survey, "The most popular sections of the paper are the features, with 80% of the students reading them. Next was entertainment (79%), sports (74%) and editorials (72%)" (Zephyr, 5/3/96, p. 2).

Producing a five-section, average 20-page newspaper required discipline and adherence to a production schedule that allowed for little time to procrastinate. Waiting until the last minute to conduct interviews or research sometimes meant a story had to be "held over" to the next issue. The habit of holding over stories was categorically frowned upon, and procrastinators became known as unreliable. No one had such a reputation during the time of my study, yet many veterans could recall past writers who had. These members left the staff after a single semester. As Troy noted, "If they can't do the work, and it's all up to you, then they can't stay. We don't have time or room for people who can't pull their own weight. It's the first thing you have to learn--to make three weeks enough time to get it all done" (4/12/96).

In order to "get it all done," the three-week cycle of production was broadly broken into three parts of the production process. The first week involved the assignment and pursuit of stories; this week ended with an editorial stance meeting during which section editors selected issues for debate, negotiated a stance that the Zephyr would take, and assigned the editorial to a writer in the group. During the second week, reporters, photographers, graphic designers, and advertising salespeople pursued their tasks of writing, creating, and selling. This was the week to complete interviews and library research and write the stories. The third and final week of this cycle was dedicated mainly to production matters: the revision and copyediting of stories and the design and layout of pages using Quark Express on Macintosh computers in the computer lab next door. The Wednesday night of this week was "Late Night on the Zephyr" and included pizza, music, and much conversational gossip as pages squeaked out of the laser printer and page composition managers painstakingly pasted the sheets to a dummy board (the newspaper-page size board with a sticky surface for lining up and arranging items). After the paste-up was complete, Thursday class time was spent in final proofing. Mr. Tillinger drove the page proofs to the printer for overnight printing and Friday distribution.

On the Friday of this third week, staff distributed the paper around the school. During classtime, the circulation manager organized the subscription labeling and mailing, the advertising sales manager assessed the financial situation, and other members of the staff submitted story ideas to the editor-in-chief for consideration in the next issue. Over the weekend, the editor selected from the stack of story ideas, and from ideas left over from past issues, and put them in a "story bank." This written bank of story ideas was organized according to the sections on the paper and used for assignments the next Monday when the whole cycle would recommence.

An important part of a member's participation in the cycle of production included talking--verbally interacting around their work. And as is true of any

profession, a specialized language allows members to communicate about their shared goals and work. Such was the case for the staff of the Zephyr. Members often talked of "leads," "attributions," "column inches," and "sidebars" in their daily interchanges around their writing and production tasks. As this terminology is central to understanding the work of the staff and is a mark of membership, figure 3.4 illustrates some of the definitions and terminology used to describe news stories.

Data Collection

My objective throughout my time with the Zephyr staff was to collect as many stories from and details of this setting from students and teachers so as to construct a rich moving picture of the interactions that created the power of words and sense of family these students talked about. The variety of naturalistic and qualitative sources of data are listed and described in Table 3.3.

As I collected and organized the data, I began reviewing tapes and documents almost immediately in order to become deeply familiar with the people and activities in this classroom. Soon after the first cycle of production, I selected 10 representative students to interview and began scheduling these conversations to take place during and after class. These interviews continued throughout the semester, as some members came to me requesting to be interviewed. I targeted five students to interview a second time--the two newcomers (Andi and Mark), two oldtimers (Troy and Dean), and one student with whom I developed an especially good rapport (Cass). I included Cass because of the relationship I was able to create with her. Heeding Dyson's (1997) advice, I "remember[ed], too, that you're looking for a teacher [student] with whom you have a rapport" (p. 22). Cass was a good teacher to me throughout the study.

Because the site for my study involved a 35 minute commute (at least) in both directions, I was able to listen to all audiotapes immediately following or preceding class sessions. This allowed me to recursively design interview protocols and to be

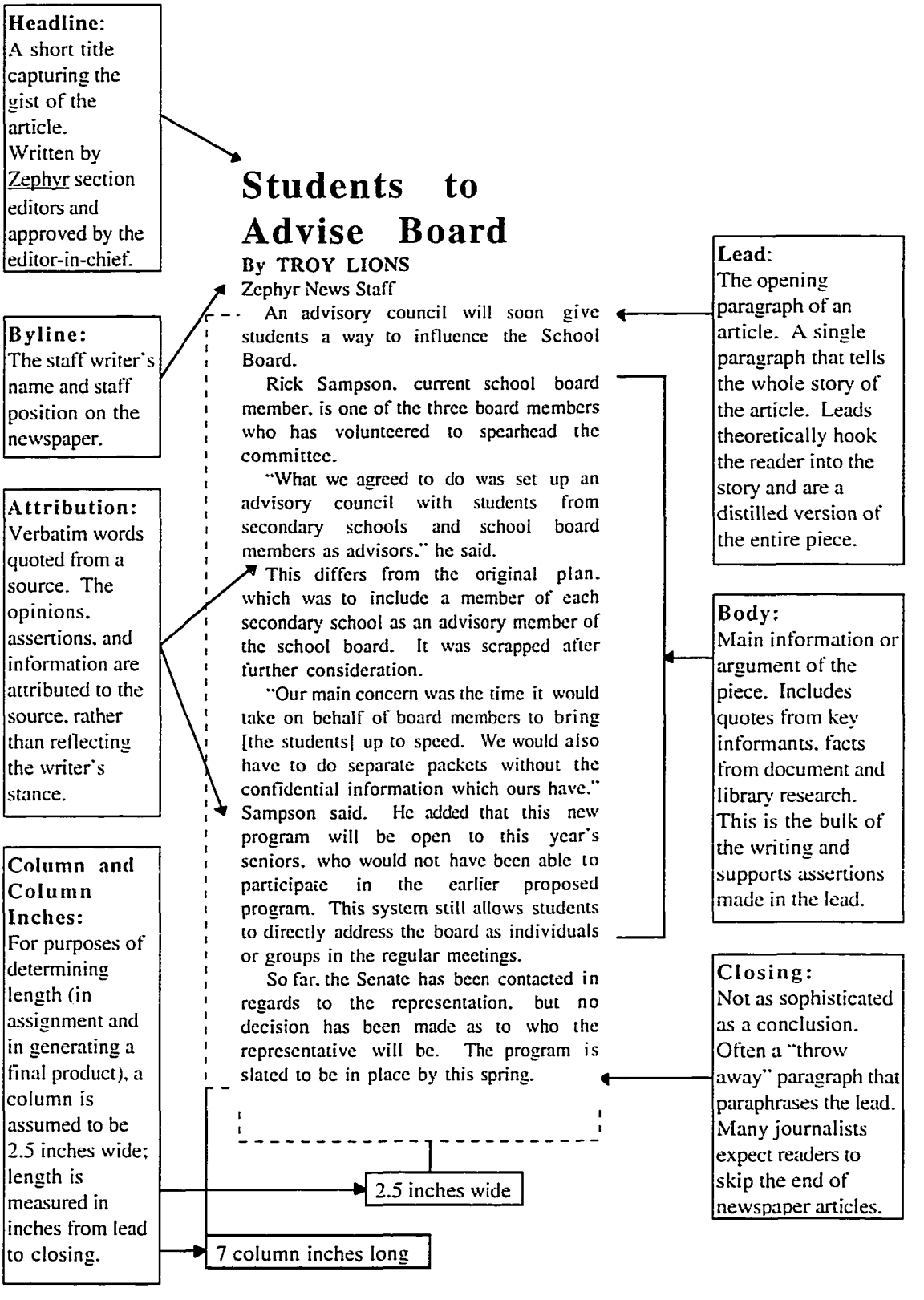


Figure 3.4 Definitions and terminology used on the Zephyr

Table 3.3

Description of Data Collection

Type of Data and Method of Collection	Description	Amount Collected
Observations Handwritten fieldnotes Audiotapes Transcription of first 15 observations and selected subsequent observations. Transcripts matched with handwritten fieldnotes.	Most often I took an observer stance toward the staff by watching and asking occasional questions. These observations sometimes turned into informal conversations, informal interviews, and group discussions which were all audiotaped.	15 consecutive 90 minute sessions (one full cycle of production) 2 X 90 minute class sessions a week for 13 weeks Total 90 minute observations = 41
Formal Interviews with Students Handwritten notes Audiotapes Transcription of all formal interviews combined with notes	I selected 10 representative students of the 23 to interview; 5 requested an interview; 5 repeated the interview at the end of the semester.	10 initial interviews 5 requested interviews 5 repeat/end of semester interviews (30-50 minutes) Total student interviews = 20
Informal Group Interviews Handwritten notes Audiotapes Selective transcription combined with handwritten fieldnotes	Occasionally an informal one-on-one conversation would lead to a group discussion/interview in which students talked among themselves in response to my questions. I took these occasions to try out interview questions as well as to gather information.	2 X Late Night sessions 2 X Friday publication day Total Group interviews = 4
Formal Interviews with the teacher Handwritten notes Audiotapes Transcription combined with handwritten notes	These questions were pre-selected and written before the interviews. Interviews lasted about an hour.	1 at beginning of semester 1 at end of semester Total Teacher interviews = 2
Other Formal Interviews Handwritten notes Audiotapes Transcription combined with notes	Dean of Students Principal English Teacher 1 English Teacher 2 3 Former Students on staff 9 Students reading but unaffiliated with the newspaper	1 X 15 minute (approx.) interview with each Total "Other" interviews = 16
Written Surveys Forms to fill out Responses coded and tabulated	I developed the survey to get a broader sense of the class. Some questions were similar to those I asked in interviews. This extended the amount of information on a few key questions, for example, "Why did you join?" and "What do you learn in a class like this?"	1 survey at the beginning of the class (20 respondents) 1 survey at the end (23 respondents) Total surveys = 43
Artifacts collected	Class notebooks (complete with notes, drafts, interview questions) Zephyr issues Applications to get into the class Applications for leadership positions Columns written by the teacher for the local paper Other pertinent news items gathered from professional presses (e.g., stories written by Zephyr staff that reflect current events and issues; stories about freedom of press issues around high school newspapers)	6 notebooks 10 issues (5 during study) 23 entrance applications 12 leadership applications 5 columns by the teacher 3 articles on student presses
Personal journal maintained on my home computer	I kept a journal almost daily during the first three weeks of data collection. Entires included reflective short essays, my initial interpretations and revised thinking. I also wrote "conceptual memos" detailing my understanding of emerging themes and patterns. I also used the journal to recorde my own emotional status and to savor surprising and memorable events.	

prepared for spontaneous one-on-one conversations during which I could ask questions that occurred to me while listening to the tapes. I listened to each complete tape before embarking on transcription.

After listening to the audiotaped interviews and classroom observations, I decided to completely transcribe all one-on-one conversations whether formal as in interviews or informal as occurred during classtime. I also chose to completely transcribe the four informal group interviews (15 to 30 minutes each) as these revealed key players, important vocabulary, and ways of interacting with each other. These complete transcripts allowed for analysis of specific language that I later marked for coding into categories and themes. Audiotapes of classroom observations allowed me to flesh out handwritten fieldnotes with verbatim conversations and to check for accuracy in my notes. For the full cycle of publication, the first 15 observations, I carefully worked back and forth between fieldnotes and audiotapes to create a faithful record of each day's 90 minute session. The nature of the class allowed for a great amount of freedom of movement and talk; thus, alone, the audiotapes provided a fairly fragmented and incomplete picture of the classroom. For the remaining audiotapes of classroom observations, I listened to these carefully a second time and selectively transcribed (Heath, 1989; 1996) one-on-one conversations and passages that seemed pertinent to the emerging and evolving themes.

The concept of selectivity in transcription is possible with the use of audiotapes. Recording events helps to counteract Erickson's (1986) warning of "the problem of premature typification" (p. 144) because the researcher can review verbatim language and conversations again and again, exploring events in depth. Heath (1989; 1996) has suggested selective transcription--making careful and grounded decisions on what and how much to transcribe--when the researcher has accumulated extended and repetitive instances of community interactions. Following the initial three-week publication

cycle, during which I did completely record and transcribe classroom interactions, I reduced my transcription output from 45 pages per class session to between 17 and 35 pages. These transcriptions were always in addition to complete fieldnotes that I carefully took at each 90 minute class session. A complete file with one day's observation, then, included: my handwritten scratch fieldnotes, the transcription or partial transcription of the class interactions (17-35 pages), and a hardcopy printout of my fleshed out fieldnotes in which I inserted in appropriate places verbatim conversations from the transcripts (40-65 pages).

The central written artifacts for my study were the final issues of the Zephyr which allowed for rhetorical and textual analyses of writing and for closer study of the writing processes used by writers on the staff. However, my collection of artifacts grew immensely when I decided to collect the class notebooks of several students I wanted to target for closer study. These notebooks included assignment sheets, notes for writing, interview questions, notes taken during interviews and research, drafts, and copyedited drafts. I collected six such notebooks and folders, made copies, and returned the work to the students the following day. In addition, I collected applications, miscellaneous handouts, and other school newspapers.

My personal journal served as a sounding board for both my ideas and emotions during data collection. It also enabled me to begin articulating patterns in the daily workings of the class and gave me a place to try out conceptual categories for further analysis. In this way, data collection was both simultaneous and recursive with initial stages of analysis and interpretation. This approach to qualitative research is well-documented in textbooks (LeCompte & Preissle, 1993; Erickson, 1986) as well as in methods sections in research literature (e.g., Athanases & Heath, 1995; Dyson, 1989; 1993).

I collected a variety of data in order to triangulate emerging themes. Data types align with the research questions and concepts that I hoped to learn about in this study.

Table 3.4 indicates how data sources connect with specific research questions and example concepts that followed from the questions. An 'x' indicates that I used that data type in my analysis of the concept and question.

A fair amount of overlap from data type to questions is to be expected in this type of research, as observations sometimes turned into informal conversations, and informal conversations sometimes took on a more interview-like structure depending on the interaction and drift of the conversation. Thus, I used different types of data to answer particular questions and address certain conceptual categories. These data were "shared" across questions to provide a complex and interwoven picture of the daily routines and activities of the Zephyr staff.

The overlap of data collection and fieldnote write-up with initial stages of analysis had the effects described by Erickson (1986):

> Write-up stimulates analytic induction and reflection on relevant theory and bodies of research literature. There is no substitute for the reflection during fieldwork that comes from time spent with the original field notes, writing them up in a more complete form, with analytic insights recorded in them (p. 144)

Thus, as I proceeded to a more careful and complete analysis of my data, I did so by first reviewing the ongoing reflections and insights I gained during data collection. In this next section, I detail the approaches and specific analytic tools I employed for the systematic analytic stage of this study.

Data Analysis

Making transparent my systematic sense-making of this group of writers is the central task of this section. The objective throughout this stage of my research has been to construct a faithful and engaging representation of the actions of the young writers for the Zephyr and to connect this representation to meaningful concepts and categories found in the literacy research literature. It is within this written representation that I can show the social worlds of these writers and the ways in which they develop literate

Table 3.4

Linking Research Questions with Types of Data

HOW DOES THE ACTIVITY OF A HIGH SCHOOL NEWSPAPER CREATE AND RECREATE SOCIAL ACTION FOR YOUTH THROUGH WRITTEN AND SPOKEN GENRES?	Observations	Student Interview	Grp Interview	Tchr Interview	Other Interview	Surveys	Artifacts
How does the wider society contribute to the creation of a high school newspaper and how does the newspaper in turn affect the wider society?							
Example Concepts							
Community	x	x	x	x	x	x	x
Social Action	x	x	x	x		x	x
Genre	x	x	x	x	x	x	x
Transformation	x	x	x	x			x
Literacy	x	x		x	x	x	x
What actions constitute the creation of a high school newspaper and how do people conform to and reform these actions in becoming members of the classroom newspaper?							
Example Concepts							
Social practices and organization	x	x	x	x			
Teamwork/Family	x	x	x	x		x	
Writing/Literacy	x	x	x	x	x	x	x
Stances	x	x	x	x			
Process	x	x		x		x	x
What are the writing processes used by individuals for personal expression and how are these operations used and adapted to create and recreate genres for social action through public expression?							
Example Concepts							
Identity	x	x	x	x		x	x
Writing process	x	x				x	x
Personal agency	x	x	x	x		x	x

identities---that is, develop a vision of literacy as a personal as well as public tool for communication and social action.

Erickson (1986) outlines a generic framework for analyzing large amounts of data collected from fieldwork. This framework begins with simultaneous data collection and initial analysis, thus making data collection a deliberate process. Following this field stage of analysis, the researcher examines "the corpus of fieldnotes, interview protocols, site documents, and perhaps audiotapes or videotapes of naturally occurring classroom events" (p. 146) for general impressions and assertions which become the basis for broad categories of meaning in the setting. Here, the researcher begins systematic searches and readings of the corpus of data for supporting and disconfirming evidence for the assertions.

Evidence in the form of data--fieldnote vignettes, interview quotes, documents--is connected to the general categories by way of "key linkages." These links organize the ways an assertion is considered--across data, across levels of social analysis (local and nonlocal social communities), and across time. As linkages are constructed around logical concepts such as time, space, consequence, and relationships, they

> are key in that [they] are of central significance for the major assertions the researcher wants to make. The key linkage is linking in that it connects up many items of data as analogous instances of the same phenomenon....In searching for key linkages the researcher is looking for patterns of generalization within the case at hand. (pp. 147-148)

These key linkages serve to connect the various levels of social interaction with the content, in this case the writing, of the community participants. Successful renderings of social science research warrant assertions with multiple types of evidence and clear, logical trains of analysis.

The following sections pull back the veil of my analytic procedures by offering concrete examples of the types of analysis and the inductive thinking that led to the chapters of findings that follow. Figure 3.5 provides an outline of my data analysis.

STAGE ONE ANALYSIS--
DISCOVERING RECURRING THEMES AND PATTERNS

Guiding Principles	Key Resources
Meaning-perspectives of participants	Erickson, 1986
Belief and doubt	Elbow, 1986
Researched world as a text	Bakhtin, 1981; 1986; Coles, 1989; Freire & Macedo, 1987; Geertz, 1973; Witte, 1992

STAGE TWO ANALYSIS--
APPLYING ACTIVITY THEORY

Guiding Principles	Key Resources
Activity	Berkenkotter & Huckin, 1993; Leontiev, 1981; Wertsch, 1981
Actions	Dyson, 1993; Heath, 1996; Kamberelis, 1995; Leontiev, 1981; Wertsch, Minick, & Arns, 1984
Tasks	Atwell, 1987; Fahnstock, 1993; Fulkerson, 1990; Graves, 1983

STAGE THREE ANALYSIS--
MICROANALYSIS OF INDIVIDUAL PARTICIPATION

Guiding Principles	Key Resources
Genre	Berkenkotter & Huckin, 1993; Kamberelis, 1995; Miller, 1984
Identity	Kamberelis, 1995; Lave & Wenger, 1991; Stake, 1994

STAGE FOUR ANALYSIS--
TEXTUAL ANALYSIS LINKED TO
RHETORICAL AND SOCIAL SENSIBILITY

Guiding Principles	Key Resources
Formal features	Bakhtin, 1981; 1986; Fowler, 1982
Rhetorical features	Fahnstock, 1993; Fulkerson, 1990
Social features	Berkenkotter & Huckin, 1993; Kamberelis, 1995; Miller, 1984

Figure 3.5 Outline of stages of analysis

Essentially, my analysis and interpretation of the data gathered for this project went through four basic stages:

1) Initial, general analysis for recurring themes and patterns;

2) application of Activity Theory to discover key linkages among three levels of analysis: activity--linked to society, actions--linked to goals, and tasks--constitutive of and linked to actions;

3) microanalysis of individual participation;

4) and textual analysis linked to rhetorical and social sensibility.

All four stages were recursive and overlapping, and all were subject to bouts of methodological believing as well as methodological doubting.

Elbow's (1986) metaphors of belief and doubt as method follow well the recursive cycles of qualitative research for, "To doubt well we learn to extricate or detach ourselves; to believe well we learn to invest or insert ourselves" (p. 264). Indeed, within Elbow's caution for balance in positive and critical interpretations of data lies a touchstone for the variety of analytic approaches used for this report. That is, when I found myself believing intensely the story I was seeing, I stopped to consciously doubt by searching for disconfirming evidence, trying another method of analysis, or considering alternative interpretations. In short, Erickson provided the outline for analysis; Elbow provided the mantra to believe and doubt.

Stage One Analysis: General Impressions and Broad Categories

I began my analysis as a reader (Bakhtin, 1981; 1986; Coles, 1989; Witte, 1992); that is, I read all the transcripts, fieldnotes, surveys, and documents with a reader's eye--looking for those elements of narrative, story, and experience that are familiar to me and taking note of those that were not. This open-book, annotating approach to my first pass through the data was decidedly fruitful, as I came to see the goings-on in the classroom as stories. Indeed, descriptions of qualitative research have often likened the scrutinized world to literary texts (Elbow, 1986, Erickson, 1986;

Freire & Macedo, 1987; Geertz, 1973; 1988) In this way, interview transcripts, fieldnote renderings, and documents are both data and texts (data-texts), and they reveal the stories of the class. In reading all of my data-texts, I found myself immersed in a world with characters, plots, settings, problems and solutions, and themes. I found myself reading the multiple stories of participation in the production of a newspaper.

When I first entered the scene, I carefully mapped the setting--the places where the action happened, where characters interacted. These places included the classroom itself, the phone room, the computer room, the cafeteria, and school classrooms and offices where staff conducted interviews. I soon concentrated my observation efforts in the classroom and computer room where students routinely interacted with one another and where they wrote, edited, and produced the paper. Within these more focused scenes, protagonists emerged, critical instances unfolded, mundane and repetitive activities as well as power structures revealed themselves. And, as with any narrative, language--both spoken and written--held the elements together with explicit and implicit themes.

I next moved back and forth between my reading-self and my writing-self. In the tradition of many professional writers (Dillard, 1989; Lamott, 1994), I used writing for discovery of thinking, patterns and theories. Laurel Richardson (1994) has connected writing as a creative process for artists with writing as an analytic process for researchers:

> I consider writing as a *method of inquiry*, as a way of finding out about yourself and your topic. Although we usually think about writing as a mode of "telling" about the social world, writing is not just a mopping-up activity at the end of a research project. Writing is also a way of "knowing"--a method of discovery and analysis. By writing in different ways, we discover new aspects of our topic and our relationship to it. (emphasis in original, p. 516)

Thus, I used writing as a tool for analysis--writing my way into my thoughts and responses to my reading. As repetition evolved from my reading of data sources, I kept notes on potential patterns in student talk, interaction, and connections with a

wider social world. I transformed these notes into "conceptual memos" that I shared with my adviser. These memos helped me begin articulating the patterns and themes, connect data from a variety of sources to these themes, and reflect on relationships to research literature. Appendix A provides an example of one such conceptual memo.

From these conceptual memos, I worked back through the data, rereading, annotating, adding and eliminating ideas. Over time, I identified six major themes that held the class together, perpetuated concepts of work and literacy, and gave me a glimpse at the meaning these students brought to and negotiated with in this classroom. I turn now to brief descriptions and examples of these six initial, guiding themes. The example vignettes and quotations are representative of the themes and are not "best case scenarios." I use them to help illustrate my logic in identifying and teasing out connections between data and theme.

Literacy

Literacy was clearly a central and guiding theme in this classroom. As students engaged in the everyday activities of producing their paper, their talk and work was filled with references, allusions, and direct mention of literate activity. While they didn't use the term "literacy" to discuss their language use, they did speak of their reading habits, their writing habits, their beliefs about the importance of writing and reading, their experiences in other classes with reading and writing, and their expectations for the role of literacy in their future lives. These subthemes around literacy appeared in their shared talk in informal conversation as well as in conversation directly linked to the writing of news articles. For example, early in February three students were sitting at desks working on articles:

> Mel, Shawn, and Cris have their notebooks open to various stages of writing. Mel is looking at some scrawled notes; Shawn is hunched over a page of prose adding to the length of his story; Cris is tapping his pencil and looking past the others. Mel reaches into her backpack and pulls out a book, Stephen King's Thinner. She asks if they've read it. Shawn says, "No, he's a hack. It's all the same." Cris looks over and says, "I used to read him. It was better--" Mel interrupts and asks if they think she should review the book for another article, "I only have

> two stories this issue. Then I can do this review and get three." Shawn notes that book reviews are boring and Thinner isn't a very good book. (2/5/96)

In this scene are issues of appropriateness of certain types of writing for the newspaper, what makes a good book, an awareness of evaluative criteria for literature, the group involvement in Mel's topic of conversation, stages of writing process--all features of literate practices. Across all data sets--written surveys, interviews, fieldnotes, and news articles--was evidence of literacy as a social practice in which members debated, challenged, worked, and learned. As this theme emerged, I searched my data-texts for such references in order to determine shared ways of understanding literacy and literate practices among members of the group.

As I built a computer file with these references to literacy, I began marking the subcategory of writing within this broader concept of literacy. Literate behavior, in contrast with "literacy skills," involves many types of language interactions: reading, writing, speaking, and listening (Heath, 1991). My interest in the relationships among genres, texts, and social practices led me to look more closely at the meaning of writing within the context of literacy in general in this classroom. As a theme, writing seemed to shift in meaning as members became more involved in the group. A striking example of this difference took place when I serendipitously observed on a day when Sperry, former editor of the Zephyr stopped by to say hello. He told me:

> I actually joined because someone interviewed me for a story. And he didn't get it quite right about me. I didn't quite agree with what he wrote. It wasn't a big mistake or anything. It made me start thinking about the power of words, the power of the paper. If you do it right, it can make a big difference. It can have great influence and change things. I decided I wanted to write where it would make that kind of difference, so applied to join. (2/13/96)

For Sperry, writing was power and it was a power to be wielded with care for it could be "right" or it could fail to represent. As an experienced writer, Sperry knew the "power of words [to] influence and change things." Others, however, had a different view of writing, focusing on more skill-based concerns, like Mike, a newcomer with a

strong background in English classes, "I had to learn the commas, the paragraphs, the attributions. Things about the format" (3/13/96). Comments like these indicated varied stances toward writing that ran from the technical aspects of writing to the social aspects of the written word. Thus, I marked direct references to writing in my file of literacy data-texts and discovered a related theme--one that showed a direct link between the act of writing and the social impact of that act--audience.

Audience

In Embracing Contraries, Elbow (1986) notes the practically universal understanding of the importance of audience for writers, and he laments the infrequency of opportunities for students to write for real audiences. Writers for the Zephyr echoed this assertion continually during the months of my study. This theme leapt from the interviews, fieldnotes, surveys, and written documents. Students individually and collectively asserted that "writing for a real audience, not just the teacher" (Cass, 3/18/96) was absolutely key to their experience. In every one of the 20 formal student interviews, the writers mentioned the critical importance of writing for real readers. Audience influenced the topics they chose to write on, the research they conducted, the language they used in writing the story. It affected the way they felt as writers and as readers. It shifted their sense of purpose in the classroom, school, and society in general. It made them aware of their peers' as well as their own biases.

Audience awareness made these students different kinds of writers from the ways they had experienced and written themselves in other classes. They mentioned how this rising awareness of the impact of their words on their peers, teachers, administrators, parents, and beyond changed the way they saw writing as a tool for social interaction, action, and reaction. For example, during a group discussion, Nic, Dano, and Amie considered the effect of their writing for the Zephyr as compared with other classes:

> I ask what they think about the role of the newspaper in school. They agree that it is to inform the students of things in the school. Nic adds,

> "It's also a chance for some of us to get a feel for what it's like to write for a paper. It's different from other classes." Dano shakes his head, "Right. It's not an essay test or anything. At least more than just the teacher reads it. You work so hard, and only the teacher reads it in those other classes." Nic says, "It's different from a paper 'out there,' but it's the same. It's an information link between the students. Everybody knows there are different groups in school....It helps the groups understand each other." Amie, Nic, and Dano talk about the weirdoes, the metal heads, and other groups. Nic says, "A lot of kids think newspaper is blow off, an easy A. and we might all get As or something, but it's much harder. We're writing for real people. the writing is for more than a class. We have deadlines. We have to be here at 11 p.m. We have to write for more than the teacher." Dano adds, "It's really different. People care about this." (2/5/96)

The growing awareness of their reading audience by "writing for real people," led to many connections among students in the class as well as out of class. That the newspaper extended their writing into a wider social world and gave meaning to their work beyond their normal school experiences with writing was a strong theme throughout the data.

Individuality and Identity

A third theme that emerged from the data focused on the individual student in the class. Individuality can be considered part of identity. Members of the Zephyr built their writerly identities through the act of writing, through feedback, through experimentation, and through interactions with others in the group. According to Kamberelis (1995), learning to write in certain genres requires a certain amount of "identity reconstruction." That is, as these students developed a niche on the newspaper and came to identify themselves as members, they sought ways to be apart from as well as a part of the group, that is, construct their own identity while identifying with a group. Examples of how students worked out their roles and identities often appeared in my fieldnotes. The following suggests the complexity of developing identities in the class:

> Troy and Lori are sitting side by side looking at the Blue and Gold, a newspaper from another school. They begin discussing whether the writers are journalists, per se. And shift to their own identities on the paper.
>
> Troy: Are you a journalist, Lori?

Lori: I don't think like a journalist through the day, so, when I'm in here I think of myself as a reporter for the paper, but when I'm dancing or in math, I don't think about myself that way. Oh, but, well, sometimes I think, "Ooo, that would be a cool story" and try to remember it for later. So maybe that makes me--

Lori goes on to say how amazing she thinks it is that the paper always comes together with so many people on it. Troy persists in asking about being a journalist.

Troy: Yes, you are [a journalist]. You get facts; you write; you publish them; You are a journalist.
Lori: [hesitating] Well, I never--
Troy: Is it the paycheck you don't get that makes the difference?
Lori: No. I guess you're right. In a literal way I am a journalist. Technically. But I don't think that way when I'm not in class. Except sometimes. No. Well. When I do other things, I think of story ideas.
Troy: But you're a journalist.
Lori: I guess. (2/7/96)

As students like Lori and Troy become more involved on the newspaper, they begin to see themselves as more than classroom participants. And working as reporters spills over into other parts of their lives as they consider stories, audience interests, and their own writing processes.

Students often talked to each other as well as to me about what it meant to them as individuals to be writers for the paper. For some, writing for the paper was part of their dream to be a writer as a career, for others it was a place to learn leadership skills. These expressions of individuality and identity appeared in many forms in my data: through topic choices for stories, in interviews about the impact of this experience on their ideas of writing in the future, in survey responses about what is learned in a class like this. As a theme, identity and individuality provided a counterbalance to the normalizing processes of becoming a member and writing within generic rules.

School

School as an institutional setting was another theme that emerged from my data-texts. Images of the school ranged from "prison" (Louis, 2/5/96) to a "place of excellent opportunity" (Troy, 3/8/96). Kamberelis (1995) and Berkenkotter and Huckin (1993) point out the import of institutional setting on genre production. These

researchers identify genre production as situated social action and note the influence of institutional settings on both surface and content features of particular types of writing. Thus, students writing for the Zephyr did so from within a powerful institutional setting--school--which provided both freedom and restriction for writing.

Almost all writers in interviews and surveys noted the opportunity to pursue topics of interest to themselves and by extension their peers. This opportunity contrasted with the usual writing they encountered in the institution. School and the types of writing usually demanded within it limited students to topics, stances, and formats generated by teachers. Writing for the Zephyr, however, allowed freedom from these institutional constraints. Alternatively, as an institutional artifact, the Zephyr had to conform to school rules for libel and appropriateness, as well as genre rules for institutional journalism. For example, swear words and patently offensive language was off limits, even though technically protected by First Amendment speech freedoms and generally eschewed by most professional publications.

How these students viewed school and their schooling experiences contributed importantly to the ways they participated and wrote for the Zephyr. The theme of school held more meaning for these students than the building or the abstraction of "institutional setting." It included issues of rules, power structures, freedom and responsibility, theories of learning, and senses of self.

Professionalism

Contrasted with the institution of school, the theme of professionalism was closely linked to awareness of audience and was compared with "real newspapers." The idea of professionalism in a school setting is an intriguing idea for those interested in formal and informal learning and for those who look for direct links between school learning and real world problem solving (Greenfield & Lave, 1982; Resnick, 1987). The Zephyr captured many features of professional journalism, for example, positions of leadership and participation, assignment and completion of articles, a real reading

audience, use of time and space linked to projects rather than class periods, consequences for writing, and a byline. In their talk and actions, writers for the Zephyr saw themselves as participating in a professional activity.

The teacher encouraged the students to see themselves as journalists involved in the production of a professional newspaper. In interviews with me as well as in dialogue with students, Mr. Tillinger made frequent reference to the professional work of the paper. Upon the occasion of receiving a letter to the editor from a teacher unhappy with a misspelled name, he gathered the class from their various work spaces:

> Mr. Tillinger has written on the board 1) ID self for interviews; 2) spell names correctly; 3) know who you are talking to. He turns and looks at the students gathered on chairs and desks. He explains that these are targets for every issue. "The first, be sure to tell people who you are, then they can decide to talk to the press, or not. As a member of the press, you have a certain authority. The second, we got a letter from an irate teacher about misspelling her name. And the third, do your homework. Know who you are talking to....We are professionals. People think we have a cool paper. We have a responsibility to uphold that credibility." (1/30/96)

The references to the students as "the press" and calling upon their "authority" and "responsibility" echoed in the words of the students themselves as they created a shared understanding of professionalism and enacted it in the classroom. As the class straddled real world and school-based writing, so too did Mr. Tillinger who played roles of both managing editor and teacher. The social as well as practical aspects of professionalism provided a thematic link to the writing and actions of participants.

Teamwork and Images of Family

The theme of teamwork, like audience, emerged from the written surveys, interview transcripts, and in the daily interactions of writers with one another. In language marked explicitly with words like "dependence" and "responsibility" and "leaning on each other," the youth writing the paper described their interactions as members of a family. In relationships marked by cooperation and consensus building, by waiting for semantic breaks to interrupt, by inviting quieter participants to talk, and

by sometimes agreeing to disagree, membership on the paper was visibly connected to relationships that built both team and family.

A successful issue of the newspaper required the efforts of all 23 staff members organizing, writing, editing, pasting up and final checking at least 36 stories. Coordinating such an effort out of diverse personalities and products demanded heavy dependence on one another to get their jobs done. Dano, Cris, Mel, and Nic discussed the family idea amongst themselves one day:

Mel: When you join you try to get to know the people.
Nic: Like Andi. She asked me to help her figure out what she is /doing today/
Mel: /good thing/ she held over all her first stories.
Cris: You stay after school. You do the work nights. Everyone depends on you. That's how you get accepted.
Mel: If they don't get accepted they aren't part of the family. This is a close group.
Dano: But you need to have a sense of humor. You need to be relaxed. Like in a family. I couldn't survive without a sense of humor. (2/5/96)

In addition to the interviews and fieldnotes, surveys indicated that the group as a whole viewed their relationships with others in the class in terms of a team or family. Part professionalism, part emotional attachment, developing the team was a central and powerful theme in my data-texts.

The six broad thematic categories presented me with opportunities for closer examination of the meaning-perspectives of participants, as each theme can be linked to research literature in writing and in social science theory. In addition, these themes had nested levels of involvement for individuals from personal, to classroom-based, to those in a wider social world. Yet, while the themes could tell a broad story of this classroom, I needed closer analysis to disentangle the various levels of participation and what those levels meant for learning and enacting literacy, audience, identity, school, professionalism, and teamwork. In this analysis, I heeded Erickson's (1986) words, "In fieldwork, one never considers a single system level in isolation from other levels: that is a basic feature of the sociocultural theory from which participant observation

methods derive" (p. 143). As a result, I turned to a more systematic examination of the levels of participation and the links among them.

Stage Two Analysis: Applying Activity Theory

The general themes identified in the first stage of analysis suggested another way of looking at the data by considering various levels of social practices within themes. For example, some social practices involved the group's involvement with a wider social world--as was evident in the themes of professionalism, literacy, and audience--such as when they wrote on social issues like homosexuality (Zephyr, 3/15/96, p. 1) or the impact of taxes spent on prisons rather than schools (Zephyr, 5/3/96, p. 3). Other expressions of this wider social involvement included legal issues of free speech and libel for writers and their group conversations around abortion, presidential elections, and school board elections. Finally, the wider social world influenced these young writers' social practices and writing based upon their own expectations for "the press" in society.

Evidence of this connection to a wider social world is not surprising given the important role newspapers have long played in societies--whether they are heavily censored as was the Soviet Union Pravda or given loose play on "facts" as in the National Enquirer. Thus, the youth involved in this newspaper production drew upon wider social expectations for newspapering as well as their own community needs for "news" appropriate for their audience. One level of social practices linked the creation of the Zephyr to a world of social, political, and legal influence that ranged beyond the walls of the school.

A second level of social practice involved the more local classroom level of interactions. Within this level could be seen the construction of a shared view of literacy within the context of the newspaper endeavor. Additionally, the day-to-day routines of production required participants to agree to certain fairly stable routines of writing and interacting. These routines were determined by both formal and informal

structures that existed in the class. The formal structure involved the characters--the statuses and roles of editor, reporter, etc.--in circumscribed patterns of behavior. The informal structure functioned simultaneously in ways that supported and sometimes disrupted the formal structure.

These local routines of interacting around story production as well as community building suggested a third level of social practice--the role of the individual performing certain tasks directed toward and sometimes subverting or altering the community goals. Social practices could be explored specifically from the point of view of the individual as s/he engaged in tasks specific to the production of writing and community. These individual ways of interacting seemed to be reflected in both the writing and participating patterns. This level of social practice involved the individual with his/her writing--the processes and intentions--and his/her personal ways of participating--as a central organizer or a peripheral observer.

These levels of social practices enabled me to see some of the ways, in broad terms, that individuals and systems work together preserving the autonomy of one another while simultaneously normalizing both. These general patterns of social practice within the classroom can be aligned with the levels of activity analysis set out by Activity Theorists (Leontiev, 1981; Wertsch, 1981). As I turned to a deeper analysis of these patterns of social practices within the six broad themes in stage two, I borrowed an analytic scaffold from Activity Theorists and filled in that scaffold with specific analytic tools borrowed from Heath (1996) and Dyson (1993).

Using Vygotsky's (1978; 1987) theory of the social construction of knowledge as a foundation, and extending ways of connecting individuals with social contexts as places for learning, Soviet Activity Theorists (Leontiev, 1981) developed the idea of Activity Systems. These systems shift the level of analysis from the individual's thinking processes to units of activity *within which* an individual thinks and acts--often in collaboration with others--to solve problems. Focusing on the activity in the present

rather than on comparisons with external, generally static, standards (Chin, 1993; Flower, 1989) of outsiders reveals how the newswriting students understood their own participation and meanings for writing as social action.

Activity systems capture individuals interacting with one another in systematic ways to achieve a common goal. When Activity is the unit of analysis, individuals learn by their participation through agreed upon actions and operations. In the following sections, I briefly define the three levels of interaction and how I analyzed my data-texts for these nested social practices.

Analyzing the First Level of Activity Settings: The Activity.

The first level of analysis focused on the *activity* which functions "to orient the subject in the world of objects" (Wertsch, Minick, & Arns, 1984, p. 154). This is the broadest view of the interactions among group members and situates the activity within the context of wider social and institutional constraints. "This first level of analysis...provides a bridge to social institutional phenomena" (p. 155). Connected with this level of human interaction is the concept of motive. Motives, constructed from historical and social sources, enable groups of people to act together, to solve problems, and to produce the materials of the activity. Motives forge the link between activities and the social context in which activities take place. A group is "motivated" to create their products in shared activity because of larger social meanings attributed to the activities and products of the group. According to Leontiev (1981):

> humans do not simply find external conditions to which they must adapt their activity. Rather, these social conditions bear with them the motives and goals of their activity, their means and modes. *In a word, society produces the activity of the individuals it forms.* (emphasis added, p. 48)

Wertsch and his colleagues (1984) take up the definition of activity in their work and note, "To specify the nature of an activity is not to specify the particular means-ends relationships that it involves but simply to identify the socioculturally defined milieu in

which it occurs" (p. 155). Thus, the activity is located within a larger social context, and its enactment is brought to fruition through socioculturally defined motives.

As defined by the Activity theorists, activity is closely aligned with an understanding of genre as social action and of critical literacy as using literate acts to affect the social world. Because activity "captures" the "norms, traditions, and ideologies" of a community, as does genre (Berkenkotter & Huckin, 1993; Kamberelis, 1995), this is an appropriate "place" to look for ways that writers create and recreate social action through their spoken and written genres--social action that connects to the wider society in which activities function.

The activity level of analysis led me to search for examples of texts, conversations, interactions, and interview responses that capture the bi-directional connections to the wider social world. That is, I searched for connections between a wider social world and a staff member's expectations for a newspaper, expectations that s/he had developed prior to and after joining the staff. And I searched for connections between a writer and his/her desire to reach out to a wider social world. I found multiple examples of both. Table 3.5 shows some example categories for this portion of the activity system analysis, a list of data sources from which the category was drawn, and a brief example of data that I chose to code as a category of "connection to wider social world." From these categories emerged patterns of connection with society--both imported and influenced. These patterns showed the ways in which the paper attempted to be like a real paper and the ways in which the paper uniquely reflected its community.

These patterns of connection were made manifest in the social practices, the actions, of the young writers. In looking for the ways that these connections were maintained and challenged, I turned to the next level of analysis in activity theory, that of the actions that constitute the activity.

Table 3.5

Coding Categories for Motivation for Activity from the Wider Social World

Categories of Influence FROM the Wider Social world	Data Sources showing this influence	Example of coded data
Purpose of newspapers	Interviews Fieldnotes Surveys	"Newspapers keep governments in check by informing the people. They are important to society." (Dean, 3/18/96)
Topics originating outside the school	Artifacts Interviews	Greenpeace actions Presidential election polls "We're just like a paper out there with news and sports and stuff." (Lori, 2/5/96)
Structural features of the newspaper	Artifacts Interviews	Sections ordered like local community paper. Langston Tribune.
Structural features of the staff	Artifacts Interviews	Organizational chart of Leadership/Editorships
Controversy and social issues	Artifacts Interviews Fieldnotes	"I read about that too. I say the gay clubs [in Salt Lake City] have every right to be there. especially if there's a Bible club. We need to do an editorial." (Louis, 3/8/96)
Categories of Influence UPON the Wider Social World	**Data Sources showing this influence**	**Example of coded data**
Motive for joining	Interviews Surveys	"I thought it would be cool to have my name in print. You know, published. I know I can make a difference when I write a good story." (Amie, 2/9/96)
Purpose of a school newspaper	Interviews Surveys Group interviews	"The school newspaper is a source of information for students. It can expose corruption too. But mainly it is to keep kids informed about the world and school." (Rusty, written survey, 5/12/96)
Perception of readers/audience	Interviews Surveys Fieldnotes	"I couldn't believe it. They read the student paper. And there were my parents' friends objecting to the story on male birth control. As if we need to be protected from information. I can't believe how close-minded they can be sometimes. And in a newspaper. What do they think it's for, anyway?" (Sera, 5/20/96)
Feedback from wider social world and issues of professionalism	Artifacts Interviews Fieldnotes	"I got a call from someone who almost wrote a letter to the editor. It seems he didn't know he was being interviewed for the newspaper. You must identify yourself, so someone can decide whether to talk to the press because we are the press." (Mr. Tillinger, 1/30/96)

Analyzing the Second Level of Activity Settings: The Actions.

The second unit of analysis Wertsch, et al. (1984) refer to is "goal-directed action." This analytic unit shifts the focus from the relationship between activity and its motives and its embeddedness within the wider society, to actions that are goal-directed and necessary to the functioning of the activity. "Human activity exists only in the form of an action or a chain of actions" (Leontiev, 1981, p. 61). Actions, then, are the social practices among participants within the activity setting that lead to successful creation and completion of the activity's motives or objects. Importantly, actions are embedded in activities and are guided by the more immediate and local goals of participants. Interestingly, the same goal-directed actions can be present within different activities. It is useful to note an example from Leontiev to illustrate the relative independence of actions and activities:

> [A]n activity and an action are genuinely different realities....One and the same action can be instrumental in realizing different activities.... Assume that I have the goal of getting to point N, and I carry it out. It is clear that this action can have completely different motives, i.e., it can realize completely different activities. The converse is also obvious: one and the same motive can give rise to different goals and, accordingly, can produce different actions...[A] kind of well-developed activity presupposes the attainment of a series of concrete goals. (p. 61)

Wertsch, et al. (1984) follow this example to its conclusion by stating, "the goal-directed action of moving from one point to another can be executed while engaged in any one of several different activities, such as labor or instruction" (p. 156). Thus, the actions of participants in activities will be goal-directed and guided by the larger motives set out by a socio-historical context.

I approached this part of my analysis by returning to the complete data-texts, particularly my fieldnotes and classroom interaction transcripts, as they best capture the social practices. I began by examining all 41 observation files for sections of fieldnotes and transcripts showing social interaction. I used several criteria to determine these data segments. First, I looked for instances of extended verbal exchanges between two

staff members or among three or more in a group. For this part of my analysis, I excluded exchanges between staff and non-staff and self-talk at the computer. Second, these exchanges needed to attend to a common theme or topic, so I excluded greetings and good-byes, and very brief, two-turn, exchanges. Frequently these topics were connected to the original six themes, as in conversations around the "boring-ness of school" or the need to write in interesting ways for their audience. And third, I separated the types of exchanges into two topical groups: topics about literacy and the production of the paper and topics around socializing. This segmenting of the 41 observation notes resulted in 73 data segments of extended verbal exchanges ranging from 15 to 250 lines of text. Conversations around production and writing accounted for about two-thirds of the exchanges (50) and socializing for about one-third (23).

I then turned my attention to those segments in which the topic or theme was directly related to writing and work on the newspaper. Thus, in considering the social practices within this group as part of generating the written genre, I concentrated on interactions around writing and newspaper production. Nearly half of the production and writing exchanges occured in formal classroom exchanges, for example, in editorial stance meetings, during the beginning-of-the-period staff announcements, and on Friday assignment sessions. The other production exchanges took place in more informal ways, for example, spontaneous peer review and editing conversations, problem-solving discussions during the production process, and Late Night critiques of other high school newspapers.

Nearly one third of the remaining extended verbal exchanges engaged students in socializing. These included topics such as planning sledding parties; complaining about sisters, brothers, and parents; and exploring the relative merits of vegetarianism.

Both types of social interactions allowed me to see the status and roles of students in the class as well as understand the recurring social practices that led to a fairly stable community over time. Through this process of identifying recurring social

practices--or recurring types of verbal exchanges--I made key linkages between the ways students interacted verbally and the types of actions that were being accomplished. Figure 3.6 illustrates an example of a key linkage between the recurring social practice of communication and problem solving through conversations about both text-making and socializing activities.

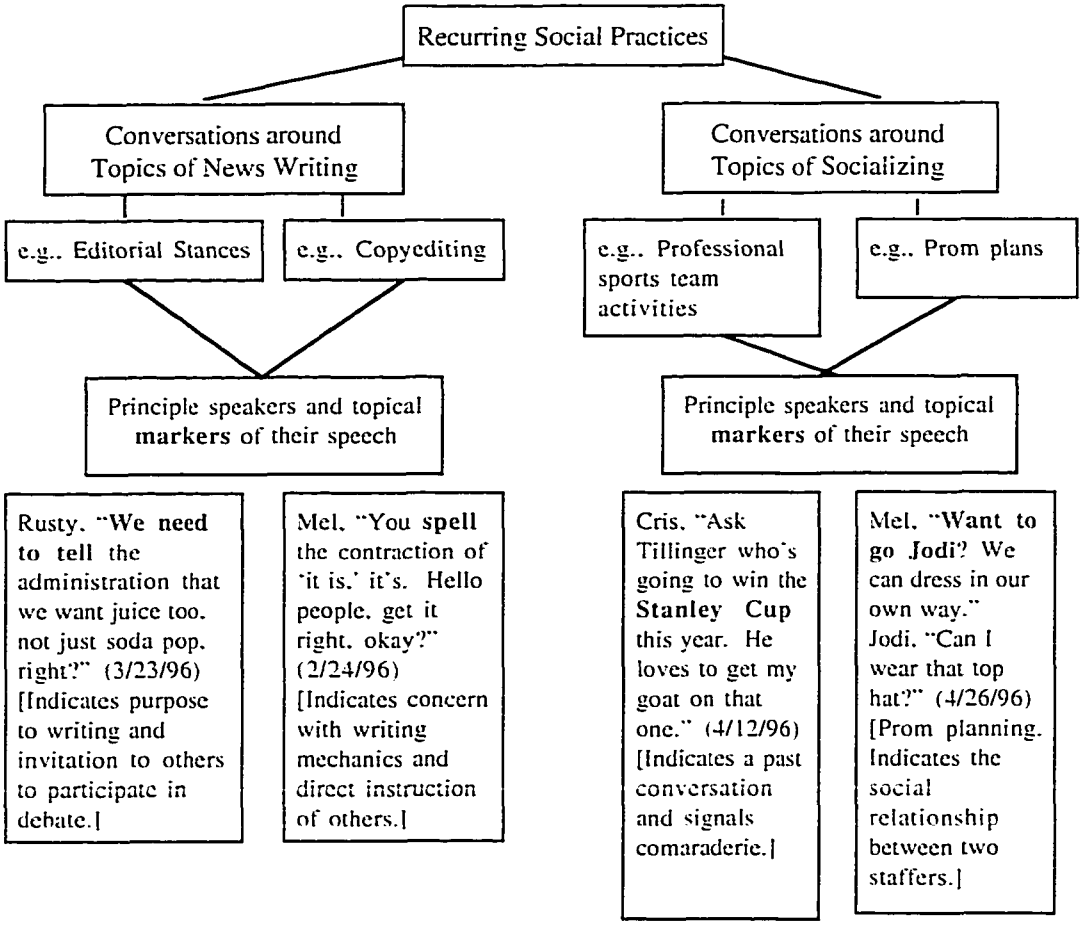


Figure 3.6 Key linkages between social practices and illustrative verbal markers

Once I had categorized certain portions of fieldnotes into these units or types of "social interactions," I proceeded to assess the quality of these interactions as a way to understand the meanings of the original six themes, for example, of "team and family" and "audience" that were clearly a part of the way students spoke of their participation and membership in the group. To perform this "assessment" of the social interactions,

I borrowed from Heath's analytic tool, "the relational analysis" (Heath, class handout, Fall, 1996).

This tool provides categories for determining the quality of interactions within groups. In order to substantiate the thematic claims of "family" and "team," for example, I examined the data transcripts for evidence of cooperation, collegiality, willingness to listen, and ability to compromise. I looked for the linguistic markers shown in table 3.6 as evidence of the group's relational characteristics (adapted from Heath, 1996). Using these analytic categories for group verbal interactions, I micro analyzed conversational data segments pertaining to both literacy and socializing. These interactions were indicators of the team and writing themes--the actions performed in the process of production of the paper. These categories created key linkages among data sources and broad thematic assertions, such as those about writing, audience, family, and school.

Once these portions of data had been coded for their relational features indicating cooperative and collective action toward the newspaper activity, I was able to see power positions and stances revealing prestige and "place" in the group with respect to literate behaviors and community building. That these positions and stances seemed to shift among speakers over time suggested that they were not static ways of interacting. That is, once a person assumed the status of "popular" or powerful in the classroom, they still took requestive and other kinds of subordinate stances toward other members of the group--a way of participating in the family. Thus, I used the relational analysis as a stepping stone toward an analysis of another social practice, taking stances toward others in the group.

This part of the analysis focused on face-to-face interactions among members of the group. As with the relational analysis, I used the conversational data segments that cohered to a topic for mulitple turns. Because stance-taking implies a type of interaction, one in which authority or obedience, for example, is practiced, this analysis

Table 3.6 Categories for Relational Analysis

Linguistic Marker	Description	Example
Content overlapping	Two people talking on topic at the same time which indicates shared purpose. "insider speech."	Cris: I hate when */people don't want to be there. They waste my time/* Dean: */there will always be people who don't want to work/*. But they take up the teacher's attention. (2/296. italics and slash marks indicate overlapped speech)
Successful interruptions	Respecting syntactic boundaries. pauses. hesitations before speaking	Cris: I'm talking about high school not middle school-- Dean: But that's the way it is now. right? Until 16? (2/2/96. interruption at end of sentence though not at the end of the full idea)
Latched speech	Finishing another's thought or sentence	Cris: Maybe they should lower it to 14 not 16. Dean. At 14 you don't know-- Mel: --how to make decisions? And you think there is something magical about 16? (2/2/96)
Facilitation of verbal involvement by another speaker	Explicit invitation to explain. clarify. elaborate. offer competing view	Josh: We know what you all think about it. But what about you. Jodi. aren't you doing a different kind of school? Jodi: I'm doing a program called a Walk About. (2/2/96)
Swarming	Rapid talk-over on same topic which indicates shared purposes and excitement.	Jodi: I get to /do what--/ Mel: /so what does that mean/ [Unintelligible as all start to take stands on mandatory schooling.] (2/2/96)
Inclusive use of "we"	Explicit attempt to gain consensus	Troy: Okay. We have to take a stand before we can write anything. It seems like Jodi and Justin have a different perspective than Cris and Dean. Could all four of you work on an editorial and an accompanying article? (2/2/96. "we" referring to the Zephyr)
Accountability	Direct call to another to pull their weight or invitation to work harder	Troy: We held over too many stories last time. That's irresponsible of us. (1/30/96) [also inclusive "we"]
Tagging	Using song snippets. free association to metaphor. alliteration. repetition indicating informal and casual conversation among or attention to friends.	Every computer is occupied as staff copyedits the issue's stories. Amie tells Lori about the Smashing Pumpkins concert she attended last weekend. Lori nods and sings a few lyrics. Jesse looks up and asks if they read that Bono. lead singer for U2. another rock group. was gay. Amie says that lots of people think rock singers are gay. Troy starts to sing. "I feel happy. I feel happy and gay." (2/13/96)
Requests for clarification or information	Asking direct questions or for repetition	Mark: How do I make one of those quote boxes and get it to fit? Dano leans over the computer Mark is working at and moves the mouse around . clicking and drawing a quote box for Mark. (3/18/96)
Direct assertions	Instructive or directive language with no hedging devices indicating authority or exercise of power	Mel: [Yelling at top volume] Hello! A lot is two words. It's two words. Get it right. We look like idiots when we do that. A lot is two words. Attention. (2/13/96)

sheds light on the both the formal and informal social organization structures that bounded the students' behavior toward one another. I borrowed this approach to analyzing social practices from Dyson (1993). In her work, she explored the stances children took toward one another while engaging in writing in the classroom. Her categories for stances (for example, appreciative, oppositional, needy) were useful as a starting point for my own stance-taking analysis. I began this analysis by borrowing from her list of categories. In the end, my list of stances was a mix of Dyson's and my own. Table 3.7 describes and illustrates the categories of stances that I used in micro analyzing the data segments of verbal exchanges.

Table 3.7

Categories of Stances Taken Toward Others

Stance category*	Example from fieldnotes and interviews
Directive	"Here, we need you to find a good quote for the quote box." (Dean to Cris, 3/18/96)
Approving	"That story was cool. I really liked it." (Jesse to Lori, 5/22/96)
Cooperative/ Willingness	"Okay, let me finish up here and I'll be right over." (Louis to Troy, 4/12/96)
Collaborative	"Let's write that one together." (Mel to Shawn, 3/18/96)
Complaining/ Protesting	"Hey, next time don't change the words without telling me. Do the spelling and punctuation, but don't change my words." (Cris to Troy, 4/22/96)
Oppositional	"I'm not signing out. That's dumb." (Andi to Nic, 2/23/96)
Helpful/ Offering expertise	"Watch this." (Dean leans over Mike and makes a box on Mike's screen while quarking the sports page, 2/23/96)
Critical	"You clown. You know you can't quote staff. That's terrible and you know it." (Jesse to Josh, 2/2/96)
Requestive	"Can you give me a hand here?" (Nic to Lori, 2/5/6)
Questioning/ Asking for info	"How do you get the person on the other end to talk?" (Shawn to Nic, 3/13;96)
Authoritative	"You've got to get those stories finished. I'm not kidding either." (Troy to class, 2/23/96)
Submissive	"I'm not the one to talk to about that. I just started." (Andi to me, 2/3/96)
Humorous	"Hey Texas, we'll just send you to Texas as our correspondent." (Mr. Tillinger to Dano, 4/12/96)
Reflective	"When I started, those stories were so shallow. Like the same old thing. Boring." (John, 5/22/96)

*Stances were not necessarily mutually exclusive.

These categories permitted me to better understand the ways students interacted with one another, ways that implied and sometimes expressed their positions of power in the classroom.

Interestingly, this analysis showed the flexible nature of taking stances. While general patterns linking particular stances to newcomers and veterans emerged, there were also many instances in which veterans took stances usually thought of as newcomer stances (for example, observing and questioning). This flexible stance-taking structure encouraged an atmosphere of mutual respect and reliance--two critical aspects of membership described by staff in interviews. Figure 3.7 provides an example of how I applied both the stance-taking and relational analyses to a portion of data (1/12/96). This interaction took place during an editorial stance session in which students discussed a snowball fight and connected it to a thrown snowball at a New York Giants football game.[5]

As I continued my second stage of analysis, determining the actions for the Activity, I sought the main actions leading to the overall activity. These analyses of social practices using Heath (1996) and Dyson (1993) were a way to understand the actions of the activity--how interactions among team members led to the six main themes, especially literacy and team building. In addition to analyzing fieldnotes for verbal interactions as part of the analysis of actions in the activity, I coded written survey responses to what students need to learn to be successful on the paper and what they will take with them when they leave. From these analyses, I was able to see how the main actions of writing and team building paralleled and supported relationships among the initial six themes. I proceeded with my activity analysis with the decision that writing and team building were co-constitutive, central actions for the activity of the

[5] In December, a New York Newsday photographer got a picture of a fan throwing a snowball onto the field. He was part of a larger snowball fight that delayed the game and earned the Giants a "crowd penalty." After publication of the photo, family and friends of the fan gave police his name and he was subsequently charged with public disturbance.

Stance-taking Analysis	**Data-text--Fieldnotes 1/12/96**	*Relational Analysis*
[Social interaction around a story/editorial assignment. 7 speaker/participants Most veteran--Nic. Troy Theme--stance on suspension and expulsion. Topic--snowball fight]	[Context: The editors are discussing taking a stand and writing an editorial on an incident in which a boy was suspended from Langston school for throwing a snowball. He subsequently returned to school during his suspension. and was expelled. The school is abuzz with the story that a kid was expelled for throwing a snowball.]	
Opposition/Critical	Mel: It seems a bit extreme to expel a kid for throwing **/a snow--/**	*Overlap speech*
	Nic: **/he wasn't/** expelled for the snowball. It was for coming to school during his suspension.	*Success interrupt*
Approving/ Authoritative/ Requestive	Troy: I think they were right-on for suspending him. **Don't you?**	*Facilitate other*
	Dean: Right. I do too. That's not the question. **Isn't that what you heard, Dano?**	*Request clarif*
	Dano: It was those kids who **/hang out over/**	
	Cris: **/hey, it's like that/ New York Giants** game where they got that guy on camera throwing the snowball	*Inapprop. interrupt Tagging/Associate*
Questioning	Dean: **What do you mean?** How's it like that?	*Request clarif*
	Cris: Well. everybody knows everybody throws snowballs. It's just this guy got caught.	
Opposition/Critical	Jesse: That's not the point. The point is more about the suspension than the expelling.	*Direct assertion*
Questioning/ Collaborative	Cris: So. what does a kid have **to do--**	*Request clarif*
	Dano: **--to get** expelled. Maybe that's where **we** take a stand?	*Latched speech Inclusive "we"*
Authoritative	Troy: If they suspend a kid. they can't come back. /What needs--/	*Success interrupt*
Humorous	Dean: /What? They/ should put the **kid in stocks** or something?	*Tagging/Associate*
	Mel: Or give them a big **red S** on their shirt. so someone **/can tell on them?/**	*Tagging/Associate (Lit. associate)*
	Jesse: **/You know/** The Scarlet Letter thing.	*Success interrupt*
	Cris: What they need is **time in the box.**	*Tagging/Associate*
Directive	Dano: Right. **A hockey penalty.**	
	Dean: **Where do we need to go with this? Who's going to write it?**	*Accountability/ Inclusive "we"*
Questioning	Troy: What are we going to write about?	*Facilitate other*
	Jesse: Right. Are **we** taking a stand to support the suspension for the /snowball--/	*Inclusive "we" Request clarif*
Directive	Troy: /The suspension/ We **need to tell** in the editorial what happened. so kids know the right story and don't get mad for the wrong thing.	*Success interrupt Direct assertion*
Cooperative	Mel: So. we support the administration?	*Request clarif*

Figure 3.7. Example of coded portion of data using stance-taking and relational analyses.

newspaper. To complete this third stage of data analysis using Activity Theory, I analyzed the actions of writing and team building for their tasks and operations.

Analyzing the Third Level of Activity Settings: The Task.

The third level of analysis is concerned with the unit of "operation." This level is yet more concrete and specific to the tasks necessary to complete the actions. Leontiev (1981) notes: "I shall label the *means* by which an action is carried out its operations" (emphasis in original, p. 63). Operations are the tasks an individual performs, under particular and local conditions, that mediate action.

> How it [the action] can be done...its operational aspect...is defined not by the goal itself but by the objective circumstances under which it is carried out.... [T]he performed action is in response to a task. The task is the goal given in certain conditions. (Wertsch, et. al, 1984, p. 156)

This micro level of analysis often focuses on individuals manipulating tools to perform certain operations necessary for the action and activity to function. While Leontiev (1981) uses examples derived from labor activity involving the use of tools for "dismembering a material object" through slicing or sawing with a knife or saw, more complex examples of "tools" might include the idea of writing as a "tool" for creating a school newspaper. It is an operation in the sense that it is performed under certain conditions (institutional, economic, personal) toward the realization of some action--writing a news story--within an activity--production of a newspaper. Understanding the tasks and their concomitant conditions will required a closer look at individuals as they composed their written stories for the paper.

I returned to my data-texts where I had identified the actions of writing and teambuilding among the coded relational and stance-taking sections as well as within the other data-texts of unabridged fieldnotes and interviews. I looked closely at the ways students spoke of their writing and teambuilding actions searching for the "parts" that made up these wholes. Table 3.8 shows the categories for these tasks within the

Table 3.8

Linking Actions to Their Constitutive Tasks

Tasks necessary for the Action of Writing	**Data Sources**	**Example**
Developing the language of the newsroom	Interviews Surveys Fieldnotes	"Did Jenne finish her biology **story**, uh, her biology paper?" (4/12/96) [distinction between "story" for news article and "paper" for academic writing.]
Topic selection	Artifacts Interviews	"I get my topics all over. And I like to check them out with my friends. Would they be interested?" (Cass, 3/18/96)
Engaging in a writing process	Artifacts Interviews Fieldnotes	"I always take a lot of notes and get quotes. Then I write a rough draft. Not always, but mostly." (Jesse, 4/12/96)
Doing research	Artifacts Interviews Fieldnotes	"I never thought of interviews as research before." (Mark, 3/13/96)
Interacting with others around stories	Interviews Fieldnotes	"Your stories are all gone over by the staff and editor. You can't be thin skinned about criticism. It's how you get better. We all help each other with our writing." (Dano, 4/12/96)
Tasks necessary for the Action of Teambuilding	**Data Sources**	**Example**
Verbal interactions	Interviews Fieldnotes	"You have to talk. If you don't talk, no one can talk back." (Dean, 3/18/96)
Tactile interactions	Fieldnotes	Jed fell into the room again today. The class cracked up. [clearly a clownish behavior that always gets a laugh] Jesse moved to help him up and he embraced her in a bear hug.
Taking responsibility	Fieldnotes Interviews Surveys	"You really have to figure out how to meet the deadlines. It gets really stressful. And so many other people are depending on you." (John, 5/22/96)
Using freedom	Interviews Fieldnotes	"This is the most amount of freedom I've ever had. And it's great. I actually work harder because I have that freedom to go out of the room, or work at home, or pick my own thing to write about." (Louis, 3/19/96)
Understanding formal structures	Interviews Fieldnotes Surveys	Formal organizational structures: reporters answer to editors; editors see that stories are finished. Formal schedule structure: three week cycles divided among production tasks.

actions and provides examples from data illustrating the links. This micro level look at the actions for the production of the newspaper contributed to my overall understanding of how this classroom system worked on a day-to-day basis. With this analysis, I felt I understood *how* the paper was actually created. But, now I needed to know more about what it meant. The six original themes and the activity analysis provided me with an in depth look at the group processes and a general feel for the overall meaning of the activity. However, my belief in this story was ready for a bout of doubt. Indeed, the analyses of the social practices of the group suggested that there were many varieties of participation. And each theme covered a range of beliefs and attitudes toward them: for example, some writers claimed to carefully consider their reader before writing, while others categorically denied being influenced by their audience. I needed a closer look at individual ways of expressing and participating to account for the ranges I saw in these first two stages.

Stage Three Analysis: Microanalysis of Individual Participation

Following an analysis of the activity, a contrary (Elbow, 1986) became evident: Activity theory may provide a good window on how to participate in the social and writing practices of the group, but it could not account for a strong theme evident in early analysis--individuality and identity. Repeatedly students noted that they chose to write and stay writing for the paper because of the personal voice they could achieve. Somehow, while students were joining the group, they were simultaneously defining themselves as individuals apart from the group. Thus, I began to look more closely at how individuals interacted with others in the class. To do so, I returned to my analyses of social practices, specifically the relational and stance analyses, and looked broadly for ways that individuals interacted and how these ways differed from one person to another. In conducting these searches, I was able to see how, for example, Troy, Sera, Dean, and other individuals created and projected their statuses in the class and how they engaged in the roles that aligned with their status. This closer look at individual

participation patterns convinced me to more carefully analyze the data and construct several mini-cases[6] to trace particular instances of membership, the individual ways of participating, and the unique approaches to identity building that seemed to run concurrently with the group processes of writing and teambuilding.

As an analytic approach, the case study has a long history in qualitative research. In his chapter, "Case Studies," Robert Stake (1994) makes a useful distinction among types and epistemological intents for the case study. He identifies three basic kinds of case study: intrinsic, instrumental, and collective. The most inductive of these is the intrinsic which ideally involves the researcher in unbiased, open-minded learning about the case. However, because my study is grounded in literacy theory (thus, I am predisposed toward looking for aspects of the class related to literacy), my choice of case study is better represented by Stake's definition of instrumental and collective. As a rationale for this stage of my analysis, he is worth quoting at some length:

> In what we may call *instrumental case study*, a particular case is examined to provide insight into an issue or for refinement of theory. the case is of secondary interest; it plays a supportive role, facilitating our understanding of something else. The case is often looked at in depth, its contexts scrutinized, its ordinary activities detailed, but because this helps us pursue the external interest....
>
> With even less interest in one particular case, researchers may study a number of cases jointly in order to inquire into the phenomenon, population, or general condition. We might call this *collective case study*. It is the study of a collective but instrumental study extended to several cases. (p. 237)

Having analyzed group processes for categories and concepts related to genre, critical literacy, and learning, I now turned my analysis toward several cases for in-depth study in order to "pursue [my] external interests."

---

[6] Stake (1994) defines "case study" as "a functioning specific...a bounded system. In the social sciences and human services, it has working parts, it probably is purposive, even having a self. It is an integrated system. The parts do not have to be working well, the purposes may be irrational, but it is a system" (p. 236). He explains that definitions of "case study" remain in some dispute that, "custom is not so strong that researchers (other than graduate students) will get into trouble by calling anything they please a case study" (p. 237).

Although my complete argument for this case study approach appears in chapter 5, here I briefly outline my justification for developing mini-cases for several students. Case studies would help me see past the system--the apparent normalizing and potentially deterministic patterns of group processes and participation. In this stage of analysis I sought to heed the contrary between the centripetal forces of the group and the centrifugal forces of individuality and identity. Elbow (1986) notes that as researchers we harbor "A hunger for coherence; yet a hunger also to be true to the natural *incoherence* of experience" (emphasis added, p. v). Case studies would allow me to see more clearly the "natural incoherence" of individuality that exists within a group.

In literacy research, Dyson (1989; 1993; 1997) has long used case studies to make important points about children's developing literate behaviors. She found that case studies allowed her to explore the individual in action with others as they co-construct and mutually construct both texts and social worlds. By conducting analyses on the activity and group processes of Zephyr staff as well as individual mini-case studies of particular writers, I hope to show the ways that these apparently conflicting and contradicting forces of membership and individuality are necessary for the life of the Zephyr. In conducting this stage of my analysis, I frequently reminded myself of what Mr. Tillinger said, "Static dies. The student paper is always changing, and that's healthy. It's alive if it changes" (1/30/96).

Reviewing the stage one and two analyses helped me think about ways to see various patterns of participation and roles. For example, I could interpret certain ways of interacting as part of the roles of newcomers, class clown, leader, loner, big mouth, etc. These ways of participating were complex and involved the current contexts-for-interacting intermingled with identities as members of peer groups, as members of families, as students, as writers, as personalities. Within this complexity, choosing the cases to study was difficult. I attended to Stake's (1994) words about selecting cases:

> On representational ground...we are optimistic that we can learn some important things from almost any case....The researcher examines various interests in the phenomenon, seeking a case of some typicality, but leaning toward those cases that seem to offer *opportunity to learn.* (emphasis in original, p. 243)

Although I could have chosen just about any staff member for these individual analyses, I selected three students who had spoken honestly (and often) with me and who willingly shared notebooks and other artifacts with me. These students were my chief informants throughout the study. Researchers doing qualitative and naturalistic work often select and rely on key informants as they survey and become part of a community scene (LeCompte & Preissle, 1993). Dyson (1997) makes the point that "inarticulate people who have poor recall and who don't get around the community much, would be poor choices [as guides or informants] to work with" (p. 22). My work was no different in this respect. Thus, based upon my rereading of the analyses pertaining to teambuilding, social practices, and actions in conjunction with the personal relationships I developed with particular students, I chose Mark, Cass, and Dean. These three allowed me to see that while "balance and variety are important; opportunity to learn is of primary importance" (Stake, 1994, p. 243).

The three members, Mark, Cass, and Dean exhibited some unique ways of participating and approaching writing, while simultaneously weaving themsevles into the fabric of the group. I chose these three as a way to understand how diverse ways of participating co-exist in an otherwise tightly knit group. For example, I selected Mark because he was a complete newcomer at the beginning of my study. As a senior beginning in mid-year, his tenure on the paper would parallel my own. I could see how newcomers were treated by veterans. And I would watch to see how Mark was pulled into the group--or possibly pushed out. In Mark's case, his personal convictions influenced his social engagement with the group as well as his writing.

The second case, Cass, was a one-year veteran at the time of my study. Her grunge dress, piercings, and bleach-dyed hair made her stand out from the more

mainstream jeans and tee-shirts. As I became familiar with her news stories, I grew interested in how her topics reflected her non-mainstream and even rebellious look as well as her loner participation patterns in the class. Though she seemed to stay on the periphery of the main social practices of the group--for instance, production nights and editorial stance meetings--her work was crucial to the paper. My interest was piqued by how a person who appears visibly and mentally distanced from the mainstream members of the class could be recognized by peers and Mr. Tillinger for her important contributions to the Zephyr.

On the other hand, my third case, Dean, epitomized the mover and shakers who make things happen in groups. He was always holding court of some sort: organizing reporters, engaging in conversations, offering encouragement, pairing people to write or quark together. As a two-year veteran, Dean had worked under two editors, had seen changes in staff, and had worked on his own writing to achieve his reputation in the class as an excellent sports reporter. As a jock, he carried some prestige with him into and out of the classroom. Dean's desire was to become the following year's editor, thus much of his attention during the time of my study was on pursuing that goal.

The three students represented a variety of membership participation patterns: they varied in time on staff, they varied in peer-group membership, they varied in their ways of participating in the group, and they varied in their ways of writing for the paper. In deciding to write mini-cases on these students I looked across all data-texts searching for their ways of enacting the writing and teambuilding themes and actions necessary for the production of the paper.

To examine their individual ways of participating in the teambuilding activities of the class, I examined their formal interviews, any informal conversations I may have had with them one-on-one in the classroom, their surveys, and all instances of their presence and participation that appeared in my fieldnotes. As I maintained computer

files of all data sources, I searched my fieldnote files for all instances of Mark, Cass, and Dean. These instances I examined and carefully cut and pasted into a new computer file, adding a brief contextual description of the portion of fieldnotes from which these data slices came. In this way, I could preserve the social context within which the interactions were taking place.

In addition, I searched the computer files of activity analysis and relational and stance-taking micro analyses for instances of Mark, Cass, and Dean's names and moved these coded interactions into their respective files. Thus, each case study file contained data pertaining to each student's social practices and participation moves as members of the class. From this analysis it was evident that social participation accounted partially for their individual ways of being members. Yet, another critical aspect of their membership centered around their writing--their beliefs about it, their experiences with it, their production of it, and their responses to it. Writing was deeply entwined within their local as well as wider social worlds which is not surprising given the central actions on the paper--teambuilding and writing.

The findings of my thematic analyses and activity analysis, plus the micro analyses of individual participation had revealed a powerful link between the content of the activity--writing--and the social practices that built-up around it. Thus, to complete my analysis of the thematic meanings of the group practices, the activity, and the individual ways of participating, I needed to turn to careful document analysis of the textual features and rhetorical techniques of the actual writing in the Zephyr.

Stage Four Analysis: Rhetorical and Textual Analysis of Writing

Because written genres are concrete and revisable artifacts of a community, their production is guided by traditional surface features in combination with evolving rhetorical features. The form, therefore, simultaneously reflects and extends the content and function of the writing. The relationship between form and function has long been the purview of rhetorical analysts (Bazerman, 1988; Berkenkotter, 1993;

Elbow, 1986; Fulkerson, 1990; Kamberelis, 1995). Therefore, in my own analysis I sought the rules that governed the surface features as well as the patterns of rhetorical engagement that were employed by individual writers of the Zephyr. In constructing these analyses of documents and written products, I could link the wider social world and the classroom social practices to the genres and texts produced by these youth. I proceeded with three categories of analysis: 1) surface features, 2) rhetorical features, and 3) social features. The first two categories would be familiar to rhetorical theorists; the final category is my own and represents an attempt to link the writing analyses to the students' meaning-perspectives of their experiences on the paper.

Analysis of Surface Features

To examine the surface features of the student work, I both quantified and qualified articles in the Zephyr. Initially, I wanted to get a feel for the newspaper genre itself, and I looked at complete issues of the paper. I identified patterns of repeated features within the newspaper, and then turned to analysis of individual writers' contributions. In this way, I could draw some comparisons between the general features of the Zephyr and the specific writing practices of individual writers.

I used categories of surface features to construct a rough set of "rules" for the appearance of the Zephyr. Table 3.9 names the surface feature categories and provides a descriptor of the counted or coded category.

After this initial analysis of the overall newspaper, I proceeded to individual writers. As part of the case study analysis, I carefully catalogued the writing in all issues done by Mark, Cass, and Dean along the surface feature categories. For each of these three writers, I created a chart of each of their stories--its location in the paper and its subgenre; its length, its use of evidence, the ratio of quote to authored prose, and use of lead techniques.

Table 3.9

Surface Features and Descriptors of Textual Analysis

Surface Feature	Descriptor
Types and amounts of subgenres	I counted and averaged the number of news stories, features, editorials, sports, and entertainment stories to get a sense of how the paper was balanced across subgenres.
Length in inches of stories	Within subgenres, I counted and averaged the length of each story type to see which stories were typically longer or shorter. (For example, sports updates averaged 8.5 inches, while features rarely were shorter than 22 inches.)
Number of paragraphs in a story	Within subgenres, I counted and averaged the number of paragraphs to explore the relative depth of detail for the story types and the amount of authored writing. (For example, stories in different subgenres varied in number of paragraphs: a ten-inch feature might have 6 paragraphs and a ten-inch club update might have as many as 12.)
Number of lines in a paragraph	Within subgenres, I counted and averaged the number of lines in a paragraph, again to explore the depth of detail in the story. (For example, editorials tended to have longer paragraphs than straight news stories.)
Techniques for leads	Question, anecdote, interesting statistic, quote, direct news; grouped by subgenre to learn the "rules" for leads.
Number of words in a story	I counted and averaged the number of words in the different subgenres as a possible way to consider vocabulary.
Quoted and authored material	Within subgenres, I counted and averaged quoted and authored material to get typical ratios. (For example, straight news stories carried a 1:3 ratio of authored to quoted material; when writers diverged from this ratio, it generally signaled conscious decisions by the writer.)
Types of evidence: Personal experience Interview information (retrievable source) Library information (retrievable source)	Counted and averaged; grouped by type of evidence within subgenres. (For example, editorials are rarely accompanied with explicit library or interview information.)

Rhetorical Features of Text

This part of the document analysis focused on the "rhetorical situation" as rhetorical theoreticians have understood it (Fahnstock, 1993; Fulkerson, 1990; Swales, 1990). The three main features for analysis are the topic, the purpose (here linked to subgenre--that is, news subgenres are for the purpose of informing; editorials are for persuasion), and the audience. Analysis of these features focused on the ways the language in the articles explored or failed to explore a topic--through the angle, the types of evidence, and the stance toward the topic. Language also revealed the author's purpose, whether it was to persuade or simply inform. Finally, language analysis showed the ways a writer attended to his/her audience through word choice, order, and selection of evidence. Thus, categories for the rhetorical analysis drew my attention to the writer's choices and uses of language. Table 3.10 provides a list and explanation of these features.

Table 3.10 Categories for Rhetorical Features of Language

Coding Category	Descriptors
Word choice	Vocabulary type and quality
Style	Formal; Informal; Idiosyncratic
Tone	Solemn; Accusatory; Irreverent; Humorous
Coherence structures	Sign-posts (use of explicit logical connectors, such as, "First...Second...." "Before." "After."); repetition; use of pronouns
Organization and use of evidence	Selection and omission of quotes and sources.
Sentence structure	Complex; simple; interconnected
Paragraph structure	Topic sentences first/last; use of detail for support.
Level of detail--Knowledge of topic	Depth of knowledge
Use of specialized language	Use of language of topic/subgenre
Literary and poetic devices	Metaphor; simile; alliteration; rhythm; rhyme; exaggeration.
Persuasive features	Ethos (authority and convention), Pathos (emotion and feeling), Logos (logic and science)
Mechanics	Correctness (spelling, punctuation, etc.)
Choice of introduction	Anecdotal; question; *in medias res* (beginning in the middle of the story, then flashing back for background); quote; statistic; the quality of the hook
Choice of closing	Dwindle; question; anecdote; frame with introduction; quote; repetition
Organization of body	Topical; temporal; pro/con; cause/effect; geographic; stringing.

I did this close-reading, rhetorical analysis for 10 stories--two from each of the five sections (news, editorial, feature, entertainment, sports). Of the two, one came from the first two issues and one came from the final two issues. I selected the articles for analysis randomly by picking a number (a playing card) between one and ten and counting from the first of the subgenre to the number I picked. For example, if there were 9 sports articles and I selected a four, I analyzed the fourth article in the section. This initial rhetorical analysis acquainted me with the features of the newspaper and with the quality of writing.

After the initial analysis, I conducted the same analysis on Mark, Cass, and Dean's articles. I analyzed one article from each issue for each writer providing me with at least five analyzed articles for each case study student. Figure 3.8 illustrates how I applied the analysis of writing techniques to student writing. Findings from these surface and rhetorical analyses helped me construct a profile of individual students as writers. These writing profiles contributed to my understanding of the thematic connections between individuals, their social worlds, and the meanings their writing held for them.

Social Features

Not wanting to lose sight of the close connection between writing and its social action in a real world, I cycled back to links between writing and social interactions. I built computer files for five students, including Cass, Mark, and Dean, out of data-texts that showed relationships among the rhetorical features of the writing, the social interaction patterns of the student in class, and their individuality and identity. Main sources of data for these links included generalizations about the kinds of topics the student tended to write on and the rhetorical risks s/he took in writing. I then compared these generalizations about an individual's writing with their discussions in interviews about their favorite and best articles as well as their explanations about the value of a

ANOTHER CRUSADE?

The First Amendment to the Constitution limits the degree government can promote a particular religion.

By TROY LIONS
Zephyr News Staff

Anecdotal opening
Historic allusion

The United States has been constantly fighting a **holy war** since the signing of the Bill of Rights and the founding of our nation.

Literary device--allusion

Extended metaphor

The **battlefield, our public schools**, have been more than **war-torn** in the last 200 years. It is free speech versus religious neutrality, **and** the **front** has moved closer to home.

Literary device--metaphor

Funnel organization: Big idea to small
National
Local community
School

With the **recent controversy** surrounding the Farview High School Holiday Program, performed at Farview High by music students, as well as the recent debate over the "Angel Tree" along a hiking trail..., has made citizens...only too aware of the hard line defense of **religious objectivity** in regards to local government.

...

According to Damian Light, the attorney for the...School District, the **Lemon test** is used to determine what is and what isn't violating the **establishment clause** in our schools.

Based on the **1971 Lemon vs. Kurtzman Supreme Court case** from which it takes its name, the test consists of three parts: if a school's action flunks any one of the three, it is found to be in violation of the **Establishment Clause**. (Zephyr, 1/26/96, p. 8)

Sentence structure--compound
Evidence--local current news
Level of detail--support
Specialized language
Evidence--interview of expert
Coherence--repetition
Specialized language
Specialized language
Evidence--library research
Specialized language
Level of detail--support

Figure 3.8 An example of rhetorical analysis of student writing.

class like this. I also looked for evidence of an individual's beliefs and attitudes toward writing in general. And I looked for connections with the individual's patterns of social engagement in the classroom. These thematic threads among a writer as an individual, his/her writing, and his/her social place in the class and wider world helped me see similarities and differences among students' ways of participating and expressing themselves. Table 3.11 provides an illustration of the triangulated data sources and thematic connections for Troy.

Table 3.11

Connecting Writing with Social Self

Category	Evidence
Beliefs about writing	Survey "Writing is a key way to get people's attention" (Troy)
Sense of self in class	Fieldnotes "Does anyone need help getting out of writers block?" (Troy, 2/23/96) [Troy takes his position as editor seriously]
Sense of self in the world	Interview "Well, I like to get things accomplished. Leadership. I'm in the class senate, too" (Troy, 4/15/96)
Topical tendencies	Documents/Interview. State and National news. Reporter. "My best article is on the First Amendment. It's important for people to know about their heritage and the constitution" (Troy, 4/15/96)
Rhetorical tendencies	Documents. Serious tone and style. Little use of literary/poetic devices. Strict adherence to who, what, when, why, where.

This final step in my rhetorical and textual analyses enabled me to see patterns linking an individual's writing choices to their social worlds in the classroom and in a wider society.

Preview of Findings

The six overarching themes: literacy, audience, individuality and identity, school, professionalism, team and family resolve themselves into the following three chapters of findings. In chapter 4, I attend to the themes of literacy, team and family, and professionalism by using my analyses to describe and interpret the membership

journey from newcomer to old-timer within the context of the Zephyr. This chapter links the literacy learning, particularly writing, to the social practices of the group as they create their written and spoken genres. I organize this chapter according to the three levels of activity (Leontiev, 1981) in order to show the shifts in the ways students understood the main actions of the activity--writing and teambuilding. Thus, I focus on my second research question: What actions constitute the creation of a high school newspaper and how do people conform to and reform these actions in becoming members of the classroom newspaper?

Chapter 5 explores the contrary between the joining processes and expressions of individuality. Here, the broad themes of literacy, individuality, and school are shown through the stories of Mark, Cass, and Dean. This chapter provides the results of the instrumental and collective case studies (Stake, 1994) that constituted stages three and four of my analysis. In this chapter, Mark, Cass, and Dean offer a window on the varieties of individual writing and participating patterns that co-exist in this group. These case studies and their attendant interpretations address my third research question: What are the writing processes used by individuals for personal expression and how are these operations used and adapted to create and recreate genres for social action through public expression? The microanalyses of individual participation coupled with the textual and rhetorical analyses reveal the tasks and operations performed at this most local level of the activity setting. Here, the focus is on the individual working the interviews, fussing with the writing, haggling with the editors, all in the effort to creating a written genre that achieves some social purpose.

Finally, chapter 6 pulls attention back away from the individual and looks out into the wider social world. This chapter attends most closely to the themes of literacy, audience, and professionalism as we see students engaging in the social work of writing. Analyses revealed that the paper's relationship with a wider social world was mutual: the Zephyr was both influenced *by* as well as had influence *on* a world beyond