Abstract

SULLIVAN, DANIEL M. Molecular and Structural Characterization of Proteins Involved in Bacterial Adaptive Responses. (Under the direction of Professor John Cavanagh.)

Bacteria are remarkable in their ability to adapt to environmental conditions that are continually in flux between growth-promoting and growth-limiting. Responses to a host of environmental situations are equally varied, ranging from the secretion of antimicrobial compounds and polymer degrading enzymes, to the up-regulation of alternative cellular developmental pathways leading to complete physiological transformation. In endospore forming bacteria this results in a metabolically inert, yet highly resistant endospore. The first study presented here deals with the NMR structural and dynamic characterization of a class of proteins in *Bacillus subtilis* known as transition-state regulators, responsible for global gene regulation during the transition from the vegetative mode of growth to the semi-quiescent stationary phase. The utilization of protein-DNA docking protocols further allows for the first description of a structural model for the interaction between these DNA-binding proteins and a cognate DNA promoter sequence. The later portions of this dissertation deal with the characterization of proteins involved in the ubiquitous bacterial signal transduction system known as the two-component signal transduction pathway. In the basic two-component signal transduction paradigm, an environmental signal is detected by a multi-domain sensor kinase that, via phosphorylation, activates a response regulator protein for its cellular role (be it DNA-binding, RNA-binding, enzymatic, etc). In the second study, a comparative modeling analysis of the predicted receiver domains the response regulators from *Vibrio vulnificus* YJ016 was performed, using the hydrophobic characteristics of the response regulator surface known to interact with the four-helix bundle of the cognate sensor

kinase as the basis for sub-classification. In the final study, a new mass spectrometric technique to detail the structural changes in proteins resulting from oxidative damage was applied to the single domain response regulator Spo0F from *B. subtilis*.

Molecular and Structural Characterization of Proteins Involved in Bacterial Adaptive Responses

by

Daniel M. Sullivan

A dissertation submitted to the Graduate Faculty
of North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Functional Genomics

Raleigh, NC

2008

APPROVED BY:

_____	_____
Chair of Advisory Committee Dr. John Cavanagh	Dr. David Bird
_____	_____
Dr. Dennis Brown	Dr. William Miller

Dedication

To the most important people in my life: my wife Lennis, and my parents Karen and Edward.

Biography

Daniel Michael Sullivan was born in Albany, New York on November 27th, 1978. His interest in science likely came from the influence of several incredible teachers during the tenure of his high school career. Out of this inspiration he pursued an undergraduate degree at Cornell University, where he received a B.S. in biology with a concentration in molecular and cell biology in the spring of 2001. While at Cornell, Daniel worked as an undergraduate research assistant for Dr. Giovannoni at the Boyce Thompson Institute for Plant Research and his interest in research science was cemented while working as an undergraduate summer researcher for Dr. Lehman at Albany Medical Center in the Department of Microbiology, Immunology, and Molecular Genetics. Daniel began his graduate studies at North Carolina State University in the fall of 2002 in the laboratory of Dr. John Cavanagh in the Department of Molecular and Structural Biochemistry.

Acknowledgements

First and foremost I would like to thank my wife, Lennis. Without your love and support I wouldn't be where I am today. Mom and Dad, words can't express my gratitude. You've been there for my mistakes as well as my accomplishments, and you didn't think I was crazy for running halfway across the country to pursue my education. I am who I am today because of everything you've done. Thank you. Dr. John Cavanagh, I would like to thank you for the opportunities you've given me over the past few years to take part in some truly exciting research. You've been a good friend and an outstanding mentor. You've pushed me harder than I thought I could push myself and you've allowed me to achieve more than I would have thought possible. Richele Thompson, thank you for your guidance, encouragement and friendship these past five and a half years. We've all relied on you, probably more than you realize. You've played the role of teacher, psychiatrist, friend, and cheering squad. We couldn't do it without you. Dr. Ben Bobay, you've been important in just about every aspect of my graduate career. If there's anyone that's done just about everything there is to do in our lab, it's you. Need help making a construct, purifying a protein, analyzing data or calculating structures? You're the one we all go to. Thank you. Dr. Douglas Kojetin, I think you've written a script to automate just about anything that can be automated, and some things that would be easier doing by hand. Thank you for all your help with the modeling project and the dynamics characterization of the transition-state regulators. The Cavanagh Lab, to everyone past and present: (Dr. David Kordys, Dr. Patrick

McLaughlin, Erin Regel, Constance Rogers, Carey Hobbs, Nicole Perry, and Sean Stowe) thank you for all the help and friendship you've given me. My Graduate Committee, I would like to thank each and every one of you for taking the time to help me during the course of my graduate school career. Dr. Mark Strauch, thank you for providing the DNA needed to conduct the transition-state regulator project. Dr. Mark Rance, thank you for all of your assistance on the transition-state regulator dynamics project. Dr. Ron Venters, thank you for your help and thank you for giving us time on your instruments. Dr. Geoffrey Mueller, thank you for all the help you provided in carrying out the transition-state regulator RDC experiments. Dr. Ken Tomer and Dr. Joshua Sharp, thank you for letting me come into your lab for my rotation. You opened my eyes to some of the outstanding work being done at NIEHS. Finally, I would like to thank everyone at the NMR Facility at NC State, particularly Dr. Hanna Gracz and Eddie Barefoot.

Table of Contents

List of Tables ix

List of Figures xi

Chapter 1. Introduction 1

Transition-state Regulators 1

Two-component Signal Transduction 11

Sporulation in *B.* subtilis 14

Aims of this work 17

Chapter 2. Residual Dipolar Coupling (RDC) Refined Structures of AbrBN, AbhN, and SpoVTN and a Structural Model of a DNA Bound Transition-state Regulator 19

Abstract 19

Introduction 19

Methods 22

Results 33

Discussion 51

Conclusions 71

Future Directions 71

Chapter 3. Sub-classification of Response Regulators Based on

Hydrophobic Surface Characteristics 73

Abstract 73

Introduction 73

Methods 81

Results 84

Discussion 95

Chapter 4. Conformational Changes in the Response Regulator

Spo0F Induced by Oxidative Damage 98

Abstract 98

Introduction 99

Methods 103

Results 105

Discussion 128

Future Directions 133

References 136

Appendices 150

Appendix A. NMR Spectroscopy 151

Appendix B. Mass Spectrometry 173

Appendix C. Circular Dichroism 183

Appendix D. Protein Expression and Purification 186

Appendix E. SpoVTN *B. subtilis* Chemical Shift Table 197

References for Appendices 211

List of Tables

Chapter 2. Residual Dipolar Coupling (RDC) Refined Structures of AbrBN, AbhN, and SpoVTN and a Structural Model of a DNA Bound Transition-state Regulator 19

Table 2.1 Definition of active and passive residues for HADDOCK modeling restraints 32

Table 2.2 NMR structural refinement statistics for AbrBN, AbhN, and SpoVTN 38

Table 2.3 AbrBN ^{15}N relaxation and dynamics parameters 43

Table 2.4 AbhN ^{15}N relaxation and dynamics parameters 44

Table 2.5 SpoVTN ^{15}N relaxation and dynamics parameters 45

Table 2.6 HADDOCK refinement statistics 47

Table 2.7 Average DNA base pair and base pair step parameters prior and post docking 50

Chapter 3. Sub-classification of Response Regulators Based on Hydrophobic Surface Characteristics 73

Table 3.1 Model statistics for *V. vulnificus* response regulator receiver domain models 86

Table 3.2 Sub-classification of response regulator receiver domains from *V. vulnificus* 88

Chapter 4. Conformational Changes in the Response Regulator

Spo0F Induced by Oxidative Damage 98

Table 4.1 Increase in the rate of oxidation as a

function of radiation dose 127

Appendix E. SpoVTN *B. subtilis* Chemical Shift Table..... 197

Table E.1 SpoVTN chemical shift values for all assigned residues 197

List of Figures

Chapter 1. Introduction..... 1

Figure 1.1 Alignment of AbrB, Abh, and SpoVT from *B. subtilis*..... 3

Figure 1.2 Schematic representation of the AbrB homotetramer..... 5

Figure 1.3 Dimeric structure of AbrBN and AbhN..... 10

Figure 1.4 The paradigm of the two-component signal transduction pathway..... 12

Figure 1.5 Stages of sporulation in *B. subtilis*..... 15

Figure 1.6 Stage 0 multicomponent phosphorelay controlling initiation of sporulation in *B. subtilis*..... 16

Chapter 2. Residual Dipolar Coupling (RDC) Refined Structures of AbrBN, AbhN, and SpoVTN and a Structural Model of a DNA Bound Transition-state Regulator..... 19

Figure 2.1 Ribbon Diagram of RDC-refined structural ensembles of AbrBN and AbhN..... 34

Figure 2.2 ^{1}H-^{15}N HSQC spectrum of 1 mM SpoVTN at 305 K..... 36

Figure 2.3 RDC-refined structure of SpoVTN..... 37

Figure 2.4 Model-free dynamic parameters 42

Figure 2.5 ^{1}H-^{15}N HSQC spectra overlay of AbrBN^{53} and oxidized AbrBN^{55} 46

Figure 2.6 HADDOCK score versus r.m.s.d for the AbrBN^{55}-tetramer in complex with *abrB8* 48

Figure 2.7 Lowest energy structure from semi-flexible Docking 49

Figure 2.8 Structural overlay of the RDC-refined structures of AbrBN, AbhN, and SpoVTN 53

Figure 2.9 Spatial arrangement of arginine residues critical in DNA-binding by AbrB, Abh, and SpoVT 54

Figure 2.10 Electrostatic surface potential plots of the RDC-refined transition-state regulators 56

Figure 2.11 Spatial arrangement of loops 1 and 2 in AbrBN, AbhN, and SpoVTN 57

Figure 2.12 GD box motif of transition-state regulators 59

Figure 2.13 Hydrogen exchange data for the “GD-Box” region of AbrBN, AbhN and SpoVTN 60

Figure 2.14 Residues experiencing motions on the micro- to millisecond timescale 64

Figure 2.15 Overlay of unbound and bound AbrBN 66

Figure 2.16 Schematic models for AbrB-DNA binding 70

Figure 2.17 AbrBN cross-linked with SuDP 72

Chapter 3. Sub-classification of Response Regulators Based on Hydrophobic Surface Characteristics 73

Figure 3.1 Structures of regulatory domains in response Regulators 75

Figure 3.2 Response regulator:four-helix bundle co-crystal structures 78

Figure 3.3 Hydrophobic scale used to visualize changes in side chain hydrophobicity 84

Figure 3.4 Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 1 – 3 90

Figure 3.5 Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 4 – 6 91

Figure 3.6 Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 7 – 9 92

Figure 3.7 Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 10 – 13 93

Figure 3.8 Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 14 – 16 94

Chapter 4. Conformational Changes in the Response Regulator Spo0F Induced by Oxidative Damage 98

Figure 4.1 CD spectra of Spo0F after varying amounts of γ-ray induced oxidative damage 107

Figure 4.2 The oxidation rate for peptide 6 – 16 109

Figure 4.3 The oxidation rate for peptide 17 – 25 110

Figure 4.4 The oxidation rate for peptide 26 – 45 111

Figure 4.5 The oxidation rate for peptide 46 – 56 112

Figure 4.6 The oxidation rate for peptide 57 – 67 113

Figure 4.7 The oxidation rate for peptide 71 – 77 114

Figure 4.8 The oxidation rate for peptide 78 – 94 116

Figure 4.9 The rate of oxidative decarboxylation for peptide 78 – 94 117

Figure 4.10 The oxidation rate for peptide 95 – 104 118

Figure 4.11 The oxidation rate for peptide 105 – 112 119

Figure 4.12 The rate of oxidative decarboxylation for peptide 105 – 112 120

Figure 4.13 The rate of oxidation for peptide 113 – 117 122

Figure 4.14 The rate of oxidation rate for peptide 118 – 122 123

Figure 4.15 Summary of structural response to oxidative damage in Spo0F 125

Figure 4.16 Increase in oxidation kinetics from 0.708 kGy to 0.944 kGy plotted on the native structure of Spo0F 129

Figure 4.17 Met81 binding pocket 132

Figure 4.18 ^{1}H-^{15}N HSQC overlay of wild-type Spo0F (black) and Spo0F Met81Leu (red) 134

Appendix A. NMR Spectroscopy 151

Figure A.1 Energy transition between α and β states 152

Figure A.2 Boltzmann distribution of nuclei in the excited (β) and lower energy (α) states at equilibrium, after a 90^{o} radiofrequency pulse, and after relaxation 154

Figure A.3 Bulk magnetization 155

Figure A.4 The FID 156

Figure A.5 One bond heteronuclear scalar couplings along the polypeptide chain which are utilized in 3D and 4D NMR experiments 158

Figure A.6 Triple resonance experiments for the assignment of $^{15}N/^{13}C$-labeled proteins 162

Figure A.7 Schematic representation of the overall process of structure determination by NMR spectroscopy 168

Appendix B. Mass Spectrometry 173

Figure B.1 Example of resolution in a mass spectrum 176

Figure B.2 Schematic diagram of a tandem Q-TOF mass spectrometer 180

Figure B.3 Example mass spectrum..... 181

Appendix C. Circular Dichroism 183

Figure C.1 Example CD spectra 185

Chapter 1: Introduction

Transition-state Regulators

The proper regulation of transcription is critical for all living cells. For the prokaryotic cell, such regulation is crucial because extracellular environments can change drastically in dramatically short periods of time. All aspects of cellular response, development, differentiation, and growth are controlled through the activity of a class of DNA-binding proteins termed transcription factors. These proteins are crucial to the ability of the bacterial cell to initiate the appropriate response to a particular environmental signal at the appropriate time. The utilization of the genetic information contained within a cell cannot be fully understood without an understanding of the structure and DNA-binding properties of these transcription factors, i.e. the structural basis for site-specific sequence recognition [1, 2]. In *Bacillus subtilis*, a soil dwelling Gram-positive bacterium, transcription factors govern the appropriate adaptive response to a myriad of environmental cues by activating and/or repressing gene expression. Sub-optimal environments, such as those resulting from changes in the availability of nutrients, temperature, osmolarity, light, pH, population density, and exposure to anti-microbial agents, lead to a shift from the exponential vegetative growth phase to the stationary phase, preparing the bacterium for survival through the expression of a host of protective genes [3].

This period of the bacterium's life, in the transition from the vegetative mode of growth to the semi-quiescent stationary phase, has been termed the "transition-state." During the

transition-state, a host of protective gene products are expressed, providing the bacterium with a window of opportunity to continue growth while assessing the quality of the environment in order to ascertain whether or not more extreme means of survival are required. The functions of the expressed protective gene products are as varied as the environmental situations that provoke such a response, ranging from expression of alternative metabolic pathways, to the secretion of toxins, proteases and variety of other degradative enzymes, the development of competence for the uptake of exogenous DNA, and the synthesis of flagella for motility, to a complete overhaul of cellular development leading to physiological transformation. In this final example, *B. subtilis* can undergo cellular differentiation leading to the development of a metabolically inert spore. The spore is the ultimate line in bacterial defense, a hardy structure that is resistant to many extremes of environmental conditions, including pH, chemicals, radiation, and heat. Spores are able to stay dormant for long periods of time, awaiting the re-establishment of favorable growth conditions, at which point the germination process allows the bacterium to resume the vegetative mode of growth.

The cellular protective functions of the transition-state are coordinated by the DNA-binding activities of a class of proteins collectively known as transition-state regulators, The role of these transcription factors is two-fold: i) they act as safeguards to prevent the unwanted expression of genes leading to adaptive responses that would be detrimental to competitive vegetative growth, and ii) they enhance the expression of genes vital to successful post-exponential growth. Though transient in laboratory situations, the transition-state is most

likely the predominant state in which *B. subtilis* exists in the natural environment. As such, the importance of this group of regulatory proteins cannot be underestimated.

The most widely studied transition-state regulators are AbrB and two paralogs with significant homology to AbrB, Abh and SpoVT, all from *B. subtilis*. An alignment of these three proteins is seen below (Figure 1.1). However, AbrB-like sequences are widely distributed among a myriad of bacterial and archeael species, with several hundred homologous sequences being found in all sequenced *Bacillus* species, as well as *Mycobacterium*, *Thermus*, *Brucella*, *Rickettsia*, *Rhodobacter*, *Haemophilus*, *Vibrio*, *Xanthomonas*, *Escherichia*, *Salmonella*, *Pseudomonas*, *Burkholderia*, *Clostridium*, *Lactobacillus*, *Geobacillus*, *Listeria*, and the *Anabaena* genera from bacteria, and *Cenarchaeum*, *Sulfolobus*, *Thermofilum*, *Hyperthermus*, *Methanospirillum*, *Halobacterium*, *Thermococcus*, and the *Pyrococcus* genera from archaea. This wide distribution among the bacterial and archaeal kingdoms only serves to underscore the importance of this class of DNA-binding proteins.

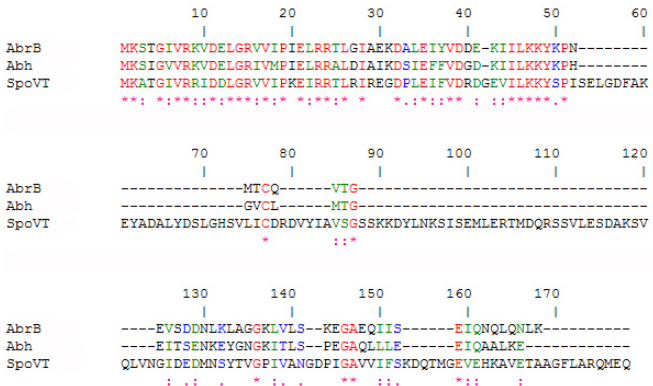


Figure 1.1. Alignment of AbrB, Abh, and SpoVT from *B. subtilis*. Identical residues are shown in red, strongly similar in green, weakly similar in blue, and dissimilar in black.

To date, the most thoroughly characterized of these proteins is the antibiotic resistance protein B, AbrB, from *B. subtilis* [4-33]. Expression of *abrB*, transcribed in a σ^A dependant manner, is subject to a level of autoregulation and is negatively regulated by the protein Spo0A, the master regulatory protein involved in the initiation of sporulation in *B. subtilis* [8, 16, 33]. The two component system leading to the activation of Spo0A, and hence the negative regulation of AbrB levels within the cell, is the subject of a later section and will not be discussed in further detail here. Once levels of activated Spo0A reach a critical level, its negative regulatory effects lower levels of AbrB within the cell, ending the transition state. AbrB is known to regulate the expression of ~60 – 80 genes nominally expressed or repressed in sub-optimal environments; this broad regulatory control is regarded as a hallmark of the transition state regulators. Originally thought to exist in solution as a homohexamer composed of 94 residues, 10.4 kDa monomers, it was subsequently shown that AbrB is in fact a homotetramer in solution [8, 25-27]. AbrB monomeric subunits are divided into two functional domains, a C-terminal multimerization domain and an N-terminal DNA-binding domain. The division of this protein into functional domains arose from a great deal of mutational analysis. While the C-terminal domain contains a predicted helix-turn-helix sequence originally thought to be important for DNA-binding, subsequent deletion-mutant analysis revealed that while the solution size of AbrB C-terminal truncation mutants decreased, the level of DNA-binding was still substantial [9]. However, mutational analysis of the N-terminal domain revealed that mutation of several arginine residues (Arg8, Arg15, Arg23 and Arg24) resulted in proteins with wild-type solution size, yet no DNA-binding activity [16, 18]. A mutation at Cys54 resulted in a protein with non-wild type solution size

and DNA-binding activity. AbrB proteins C-terminally truncated at position 54 (thus consisting of N-terminal residues 1-53) exist in solution as a stable dimer. Yet these dimeric DNA-binding domains showed a significant decrease in the affinity for DNA target sequences. However, protein constructs consisting of residues 1-55 and containing a disulfide bond between the cysteines at position 54 form a dimer of dimers (tetramer) in solution. This higher order oligomerization restored DNA-binding affinity to near wild-type levels in the absence of the C-terminal domain [18, 20]. As such, the N-terminal residues 1-53 of AbrB are designated a dimerization and DNA-binding domain and residues 54-94 of the C-terminus are designated as a dimerization domain responsible for formation of the higher order tetramer. The putative organization of AbrB subunits, as well as their role in multimerization of the full length protein is shown below in Figure 1.2.

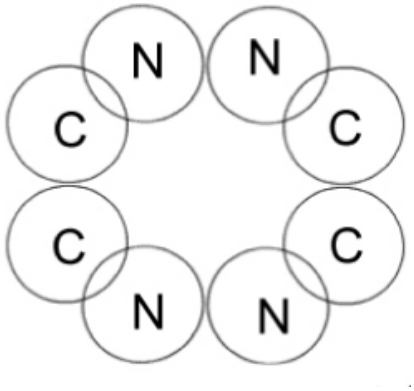


Figure 1.2. Schematic representation of the AbrB homotetramer. Monomeric subunits are shown as two overlapping circles (labeled N and C). Multimerization interfaces are noted by circles contacting edge-on.

Despite the high number of known *in vivo* promoter sequences with which AbrB has been

shown to interact, no consensus base sequence is evident. High-affinity sites of AbrB-DNA interaction selected using *in vitro* methods have been shown to converge to a loose pseudo consensus sequence, TGGNA-5bp-TGGNA. However, sequences resembling this pseudo consensus are rarely found in chromosomal binding sites [5, 7, 9, 12, 14, 15]. It is widely recognized that AbrB does not interact with every nucleotide sequence and it has been shown to interact with a subset of DNA-binding sequences with high affinity, with Kd's in the low nM range [12]. Methylation and hydroxy-radical footprinting experiments have shown that DNA-binding occurs on one face of the DNA molecule and DNase I footprinting analysis has revealed a variable amount of protection, ranging from 30 to 125 bases [7, 9]. Finally, although only a pseudo-consensus sequence is apparent, cognate promoter regions are high in A+T content, indicating that some level of conformational flexibility is critical for DNA-recognition. This final point is of interest in that it indicates that the ability of the DNA target sequences to adopt a variable subset of three-dimensional structures is important for recognition and binding by AbrB [17, 34]. Taken together, a wealth of genetic and biochemical analysis suggests that the binding properties of AbrB are three-fold: i) non-specific interactions arising from the recognition of general DNA features, ii) limited promiscuity allowing for interactions with a subset of structurally related DNA sequences, and iii) high affinity interactions with specific DNA sequences [35].

While a great deal of information is known about AbrB, the homologous transition-state regulators Abh and SpoVT remain less well characterized. *abh* (antibiotic resistance protein h) is transcribed from a σ^X dependant promoter and has been shown to regulate an array of

antimicrobial responses during the transition state in *B. subtilis* [36-38]. As is the case for AbrB, Abh is thought to exist functionally as a homotetramer in solution, composed of 92 residue, 10.2 kDa monomers. Abh is also divided into two functional domains, an N-terminal DNA-binding domain made up of the first 54 residues, and a C-terminal multimerization domain made up of residues 55-92. As is the case for the N-terminal domain of AbrB, the N-terminal domain of Abh exists as a stable dimer in solution. The DNA-binding properties of Abh have been shown to be significantly different from those of AbrB. In contrast to the sharply defined DNase footprinting assays of AbrB, Abh seems to have a more relaxed binding specificity than AbrB, extending into adjacent DNA regions in a sequence independent fashion. Abh has been shown to have a pH-dependant binding profile, centered around pH 7, to a common DNA target. Abh has also been shown to bind to the promoter regions of genes which are also regulated by AbrB as well as to several of the high affinity AbrB-binding DNA aptomers, yet DNase footprint analysis has revealed that Abh and AbrB bind in different, yet partially overlapping manner to DNA sequences [35].

SpoVT (sporulation stage V protein T), the third AbrB paralog in *B. subtilis*, has been shown to regulate a number of genes in the σ^G operon. *spoVT* is transcribed at two times during cellular growth, first as the result of read-through transcription from the upstream *mfd* gene during vegetative growth, and 2.5 hours after the onset of the stationary phase in a σ^G dependant manner. However, it is only during the later σ^G dependant transcription that the reading frame coding for SpoVT is transcribed [39]. Though not active during the vegetative growth mode or during the transition to the stationary phase as is the case for AbrB and Abh,

SpoVT is classified as a transition state regulator due primarily to its sequence identity to AbrB and Abh as well as its DNA-binding properties. SpoVT acts both positively and negatively in controlling σ^G-dependant gene expression and is absolutely required for proper spore formation, as deletion of *spoVT* results in spores exhibiting defective spore coats with reduced resistance properties [39-43]. As is the case for AbrB, no consensus sequence is obvious from the promoter regions of genes regulated by SpoVT [43]. In regards to organization SpoVT, a 178 residue 19.6 kDa protein, is divided into two functional domains as are AbrB and Abh, an N-terminal DNA-binding domain made up of residues 1-55, and a C-terminal multimerization domain comprised of residues 56-178. The N-terminal domain has been shown to exist as a dimer in solution, while the full length protein is thought to exist as either a dimer or as a tetramer [43, 44].

As no consensus DNA binding sequences for AbrB, Abh, and SpoVT have been recognized, and no sequence has been found that is recognized by all three proteins in the same manner, a central issue arises: how does each transition-state regulator select for the correct subset of binding sites to carry out the proper biological function? Further complication the issue is that the C-terminal domains of AbrB and Abh have been shown to be independent multimerization domains with little, if any, role in conferring DNA-binding specificity. As such, observed differences in DNA-binding properties of these proteins must reside in the limited difference in the amino acid composition of their N-terminal domains. In an effort to fully investigate the structural similarities and/or differences that allow these proteins to

recognize a disparate collection of cognate promoter sequences, structural studies have been carried out on the N-terminal DNA-binding domains of AbrB (AbrBN) and Abh (AbhN) [28, 35]. Given the high degree of sequence identity in the N-terminal regions of these proteins, it is not surprising that these proteins share a great deal of structural similarity, witnessed by a r.m.s. deviation of only 1.73 Å. Both proteins are dimers, each monomeric subunit being composed of four anti-parallel β-strands and a single α-helix arranged in a domain swap fold (Figure 1.3A-D). In this fold β-strands 1 and 2 from monomer 1 interact with β-strands 3 and 4 from monomer 2. In both structures, an extensive four-stranded β-scaffold comprised of β-strands 3 and 4 from both monomers is present and an extensive dimerization interface is seen between β-strands 4 from both monomers. The α-helices further displayed significant structural similarities, including length and angle/pitch with respect to β-strand 2. Given the that these structural similarities would have implications for a similar mode of DNA-recognition and binding, the arginine residues noted above as being critical for the ability of AbrB to bind DNA (Arg8, Arg15, Arg23, and Arg24), were all found to occupy structural conserved orientations on the same face of the dimeric structure (Figure 1.3E and F). Given the above similarities and their proposed roles in general mechanism of DNA-recognition and binding, several regions of structural and/or electrostatic differences were noted, primarily in the positioning of loops 1 and 2. This positional shift, as well as electrostatic differences in loop 2 and the region surrounding Arg-8, can be expected to play a role in the specific DNA-binding properties of the individual transition-state regulators.

The first portion of this dissertation continues the above structural analysis of the transition-

state regulators, extended to include the determination of the NMR solution structure of the N-terminal DNA-binding domain of SpoVT, as well as structural refinement of all three structures. The dynamic motions of the transition-state regulators are also characterized, detailing similarities and differences with potential functional implications.

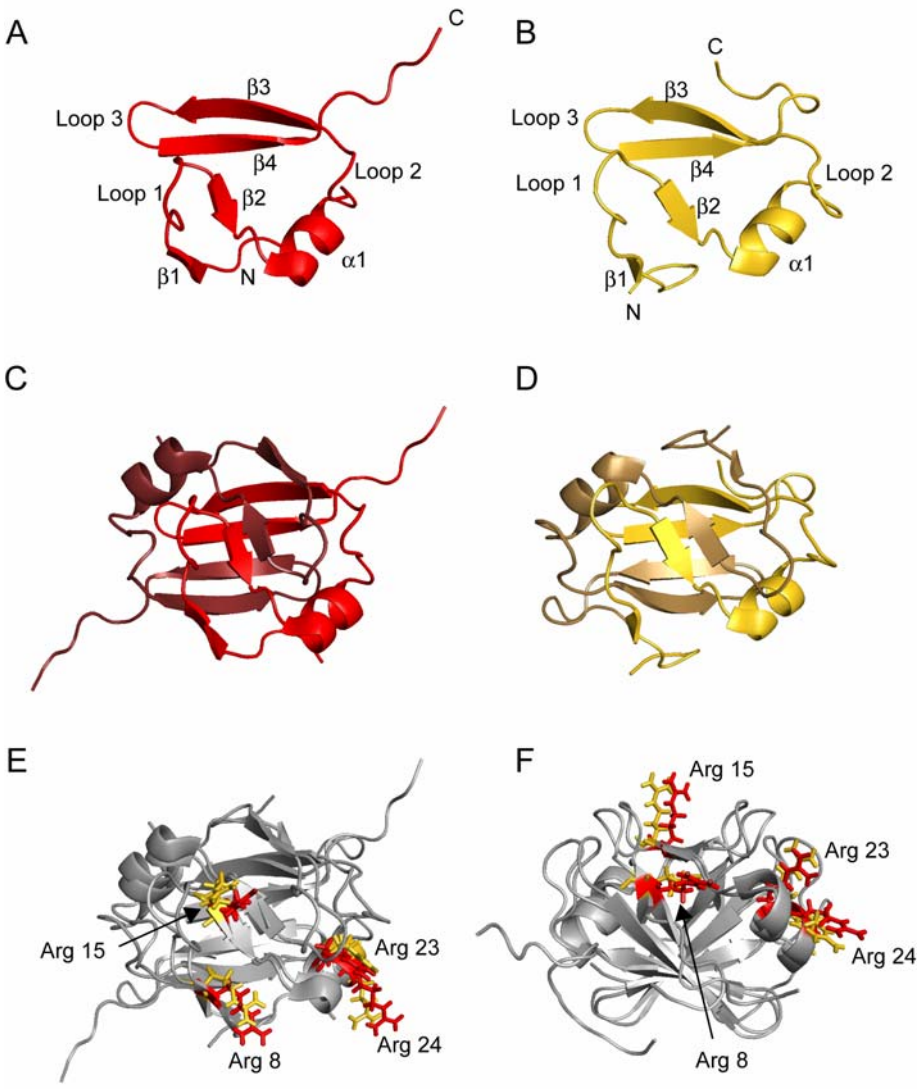


Figure 1.3. Dimeric structure of AbrBN and AbhN. *A* and *B* Monomeric subunits of AbrBN and AbhN, respectively, with secondary structure elements labeled; *C* and *D* Dimers of AbrBN and AbhN, respectively, with monomeric subunits denoted by light and dark coloration; E and F Arginine residues key for DNA-binding in AbrB and Abh are depicted from one monomer for clarity (residues are red for AbrBN and gold for AbhN)

Two-component Signal Transduction

In order to adapt to a vast array of environmental stresses, bacteria must posses the capability to sense specific signals within their environment and translate this signal into coordinated response. This capability is provided through the use of the elegantly simple two-component signaling system, present in both Gram-positive and Gram-negative bacteria, as well as archaea, fungi, and unicellular eukaryotes [45-48]. While the process controlled by the two-component signal transduction system may vary tremendously, from DNA- and RNA-binding to enzymatic processes, all two-component systems are composed of similar core domains: i) a sensor histidine kinase, and ii) the response regulator [45, 49, 50]. Specific environmental signals are detected through highly variable N-terminal sensing domains of the sensor histidine kinase, resulting in an ATP-dependant autophosphorylation event at a conserved histidine residue on the sensor histidine kinase. This phosphorylation event then allows the kinase to be recognized by its cognate response regulator binding partner. The phosphate group is then transferred to a conserved negatively charged residue, predominantly an aspartic acid, on the receiver domain of the response regulator. The transfer of this phosphate group to the receiver domain of the response regulator activates the response regulator to carry out the molecular activity of the output domain. Due to the nature of the residues that receive the phosphate group, the two-component signal transduction pathway is also referred to as the histidyl-aspartyl (His – Asp) phosphorelay [51]. The domain structure, and flow of information, in the ubiquitous two-component signal transduction pathway is diagrammed below in Figure 1.4.

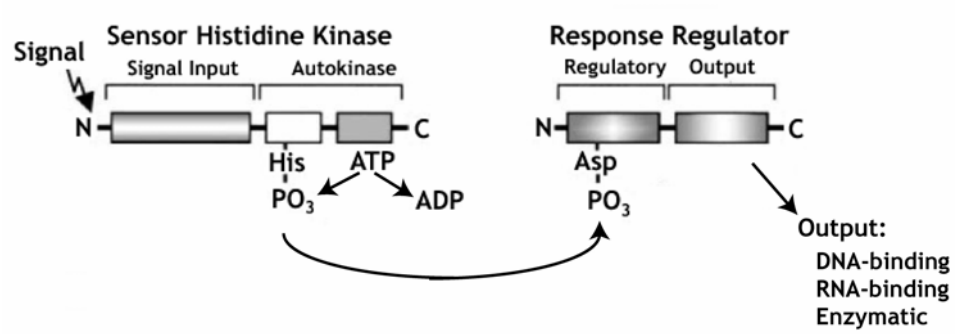


Figure 1.4. The paradigm of the two-component signal transduction pathway. Upon receiving an environmental signal, an autophosphorylation event in the autokinase domain leads to phosphorylation at a conserved histidine. This phosphoryl group is transferred to the regulatory domain of the response regulator, activating the C-terminal output domain for its cellular activity.

Sensor histidine kinases are predominantly integral membrane proteins, though cytosolic examples are not uncommon [52]. Sensor histidine kinases are primarily divided into two domains, a variable sensing domain, and the conserved catalytic core. Environmental signals are detected via the N-terminal sensing domain, leading to the phosphorylation event in the catalytic core, which is comprised of an ATP-binding catalytic domain as well as a dimerization domain. The dimerization domain forms a four-helix bundle, allowing the phosphorylation event to occur in *trans*, where the catalytic domain of one subunit in the homodimer phosphorylates the conserved histidine in the second subunit [47]. The sensor domain, residing in the N-terminus of the protein, shows a great deal of sequence and structural variation, allowing for specificity in signal recognition across various two-component signaling pathways. This N-terminal domain is connected to the C-terminal catalytic core domain via a linker domain of variable length, ranging from 40 – 180 residues. While little is understood of this linker domain, it is known that they are critical for signal

transduction [47, 53].

The second component of the two-component signal transduction pathway is the response regulator [47]. These proteins are generally composed of two domains, an N-terminal regulatory (or receiver) domain, and a C-terminal effector domain, though other possibilities are possible, such as single domain response regulators containing a lone regulatory domain as well as three domain proteins. The N-terminal regulatory domain is approximately 125 residues forming an α/β domain with a central β-sheet core made up of five parallel β-strands surrounded by five α-helices, a fold common to all response regulators [47, 51]. This domain contains an active site aspartic acid phosphoryl binding pocket that is made up of three conserved aspartic acids, one of which is the recipient of the phosphate group from the sensor histidine kinase, as well two conserved threonine and lysine residues. These residues are involved in coordinated activities including binding of the divalent metal ion required for phosphorylation, stabilization of the phosphoryl state, and autocatalysis mechanisms [54-57]. The regulatory domain exerts a negative, inhibitory function upon the C-terminal output domain. Upon phosphorylation of the N-terminal regulatory domain, this inhibitory effect is abolished, allowing the C-terminal output domain to perform its function. The function of this domain is generally that of DNA-binding leading to transcriptional activation or repression of genes in response to a particular environmental signal. However, some C-terminal output domains have other functions, including RNA-binding, enzymatic, as well as protein-binding [58].

Most two-component signal transduction systems follow the simple paradigm outlined above. However, more complex variations of the two-component signal transduction systems exist, which integrate multiple two-component pairs into the pathway, allowing for further regulatory checkpoints along the signal circuit. The more complex two-component systems are referred to as a multi-component phosphorelay, in which an environmental signal leads to the autophosphorylation of the conserved histidine of the sensor histidine kinase, which is transferred to the conserved aspartic acid of the regulatory domain of a single domain response regulator. The phosphate is subsequently transferred to the histidine of the histidine phosphotransfer domain (Hpt) of a phosphotransferase. In the final step, the phosphate is transferred from the histidine of the phosphotransferase to the conserved aspartic acid of the regulatory domain of a second response regulator. In this manner, the multicomponent phosphorelay is termed a His – Asp – His – Asp phosphorelay [57]. The extensively studied sporulation initiation pathway of *B. subtilis* is one example of such a multicomponent phosphorelay. In this signal transduction pathway, phosphoryl transfer ultimately serves to activate the master switch of sporulation, Spo0A [59-61].

Sporulation in *B. subtilis*

In the transition from vegetative growth to the stationary phase, *B. subtilis* may undergo a process of alternative cellular development known as sporulation in response to a host of unfavorable environmental cues, including nutrient deprivation and increased cell density [3, 62-65]. The process of sporulation is comprised of seven stages during which a vegetative cell is transformed into a large mother cell within which is the developing forespore [66, 67].

Although both the mother cell and the forespore have an identical genetic complement, a series of cell specific alternative RNA polymerase sigma factors direct gene expression resulting in divergent developmental fates [62, 68]. Once the forespore is engulfed in a protective proteinacous coat and is dehydrated, the mother cell is lysed, releasing the mature, metabolically dormant endospore. This developmental pathway, known to involve upwards of 215 proteins, is shown below in Figure 1.5. The endospore is extremely resistant to heat, low pH, chemicals, sunlight, and drugs. Sporulation is an energy intensive terminal process not to be entered spuriously. As such, the initiation of sporulation is under the tight regulation of a multicomponent phosphorelay pathway (depicted below in Figure 1.6).

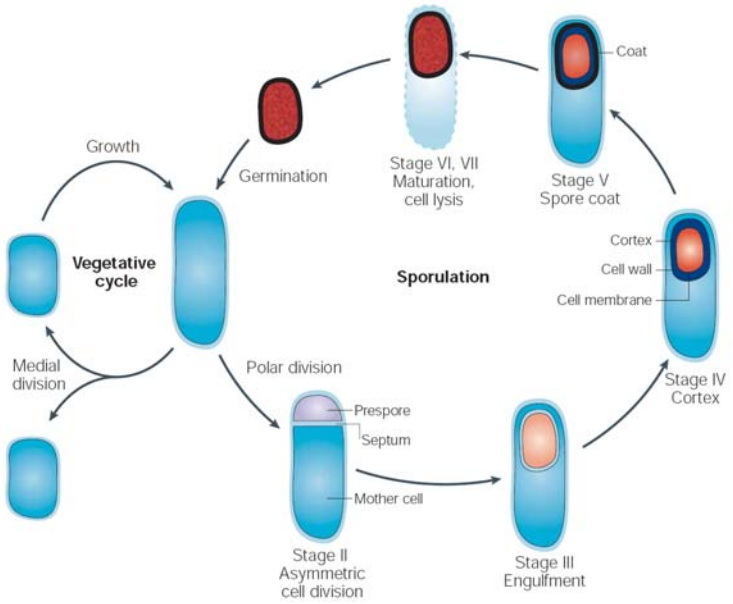


Figure 1.5. Stages of sporulation in *B. subtilis*. Cells that leave the vegetative state form a pair of partially replicated chromosomes in an axial filament in Stage I. In Stage II, asymmetric cell division separates the mother cell from the developing pre-spore. In Stage III the pre-spore is engulfed in a phagocyte-like process, creating a free protoplast within the mother cell. In Stages IV and V a loose peptidoglycan cortex and a proteinacous spore coat are assembled on the surface of the forespore. In Stage VI maturation of the spore allows the spore to acquire full resistance properties. At Stage VII sporulation is completed with the lysis of the mother cell.

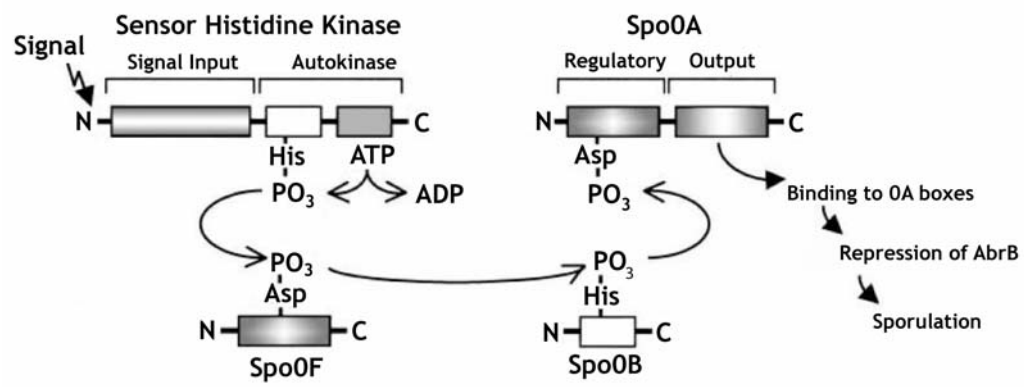


Figure 1.6. Stage 0 multicomponent phosphorelay controlling initiation of sporulation in *B. subtilis*. One of several sensor histidine kinases initiates this phosphorelay. Once a signal is received, these kinases autophosphorylates on a conserved histidine. The phosphoryl group is subsequently transferred to the Asp pocket of Spo0F, then to a conserved histidine in this Hpt domain of Spo0B, and then in the final step to the Asp pocket of the response regulator Spo0A. This releases the inhibitory effect of the regulatory domain, activating the DNA-binding activity of the C-terminal output domain.

This pathway is initiated by one of five sensor histidine kinases: KinA, KinB, KinC, KinD, and KinD [61, 69, 70]. KinA has been shown to be the primary kinase involved in the *in vitro* initiation of the phosphorelay. When activated by yet unidentified environmental signals, KinA autophosphorylates on a conserved histidine in an ATP-dependent manner. This phosphoryl group is then transferred to an aspartic acid in the single domain response regulator Spo0F, which is then transferred to the phosphotransferase Spo0B. In the final step the phosphoryl group is transferred from Spo0B to a conserved aspartic acid in the regulatory domain of the response regulator Spo0A. This phosphorylation event leads to the activation of Spo0A, a transcription factor binding to the conserved DNA sequence TGNCGAA, otherwise known as a 'Spo0A box', found in the promoter regions in genes of the Spo0A regulon. Spo0A positively regulates not only its own regulation, but also genes required for

later stages of sporulation, including genes required for axial filament formation, asymmetric cell division, and sigma factors σ^E and σ^F. Spo0A negatively regulates a number of genes as well, including the transition state regulator *abrB*, *spo0E*, and *spo0H* [71][72, 73]. As mentioned previously, one of the benefits of signal transduction via a multicomponent phosphorelay pathway is the introduction of further regulatory checkpoints. In the phosphorelay controlling the initiation of sporulation, multiple proteins exist to dephosphorylate the response regulators Spo0F and Spo0A, thus terminating signal transduction. Spo0F may be dephosphorylated by one of three phosphatases (RapA, RapB, and RapE), while Spo0A can be dephosphorylated by Spo0E, YnzD, and YisI [74-76]. These various proteins afford a multitude of mechanisms by which the initiation of sporulation may be controlled, as the decision to commit to this strategy is of paramount importance; once asymmetric septum dividing the mother cell from the developing forespore has formed, the bacterium is committed to the process.

Aims of this work

The work presented in this dissertation describes the studies performed on three separate projects. Chapter 2 describes the structural and dynamic characterization of the N-terminal DNA-binding domains of the transition-state regulators AbrB, Abh, and SpoVT from *B. subtilis* as well as the elucidation of a mode of interaction between the N-terminal DNA-binding domain of AbrB and a cognate promoter sequence using NMR spectroscopy. Chapter 3 describes a comparative modeling study of all putative response regulator receiver domains from the organism *V. vulnificus*, to provide insight into the inter-relatedness of this

family of proteins. Finally, Chapter 4 describes the development of a mass spectrometric based methodology to study the structural effects of protein oxidative damage using the well characterized single domain response regulator, Spo0F, from *B. subtilis* as a model system.

Chapter 2: Residual Dipolar Coupling (RDC) Refined Structures of AbrBN, AbhN, and SpoVTN and a Structural Model of a DNA Bound Transition-state Regulator

Abstract

Understanding the nature of DNA recognition and binding by the AbrB-like family of transcriptional regulators is of great interest as these proteins are crucial to the ability of bacteria to elicit the appropriate adaptive response to a particular environment. Although a great deal of characterization has been accomplished at the genetic level, the general and specific mechanisms of binding remain elusive. Here we present the high-resolution RDC-refined NMR solution structures and dynamic properties of the N-terminal DNA-binding domains of the *Bacillus subtilis* transition-state regulators AbrB, Abh, and SpoVT. We also combine DNase I footprinting, DNA methylation, gel shift assays, mutagenic and NMR studies to generate a structural model of the complex between AbrBN and a cognate promoter, *abrB8*. This analysis provides the first insight into the specific nature of the transition-state regulator-DNA interaction and represents the first model for any protein in the 'AbrB-like' class bound to a DNA target.

Introduction

Bacterial survival and adaptability is contingent upon the initiation of the appropriate response to a particular environmental stimulus at the appropriate time. These environmental stimuli are varied, including changes in temperature, osmolarity, pH, and cell density [3, 77].

Exposure to the suboptimal environments resulting from these types of environmental changes leads to a shift from the vegetative growth mode to the stationary phase along with a concomitant shift in transcriptional regulation, leading to an up-regulation in protective genes. Given the vast array of environmental stimuli, bacterial responses are equally varied, ranging from the secretion of toxins and polymer-degrading enzymes, and the synthesis of antibiotics to cellular differentiation. The end result of this cellular differentiation pathway is the bacterial endospore, a metabolically dormant structure highly resistant to heat, sunlight, chemicals, and drugs. In *Bacillus subtilis* the responses to such diverse environmental stimuli are regulated by a family of transcription factors known as the transition-state regulators [10].

Numerous 'AbrB-like' transition-state regulator proteins have been identified in a wide range of organisms, including species from the *Bacillus*, *Clostridium*, *Geobacillus*, *Listeria*, *Staphylococcus*, and *Streptococcus* genera [10, 23, 36, 39, 43, 78-80]. The most widely studied of the 'AbrB-like' proteins has been *B. subtilis* AbrB, a homotetramer composed of identical 94-residue monomeric subunits [7, 8, 12-17, 20, 22, 25, 27, 28, 34, 38, 56]. AbrB regulates the transcription of more than 60 different genes. Yet despite the number of known DNA binding targets, only a loose pseudo consensus nucleotide sequence is apparent – TGGNA-5bp-TGGNA. It is recognized that AbrB does not interact with every nucleotide sequence and it has been shown to interact with a subset of DNA targets with high affinity (with Kd's in the low nM range) [17, 27]. The DNA-binding properties of AbrB have been described as three-fold: i) non-specific interactions arising from the recognition of general

DNA features; ii) limited promiscuity allowing for interactions with a subset of structurally related DNA sequences; iii) high affinity interactions with specific DNA sequences [35]. Understanding the three-fold nature of this DNA binding activity can only be achieved by comparative structural studies carried out on structurally homologous proteins that display different binding properties while at the same time retaining general recognition characteristics.

The first such study between 'AbrB-like' transition-state regulators has recently been carried out. A comparison of the NMR structures of the dimeric N-terminal domains of AbrB (AbrBN) and Abh (AbhN) highlighted a number of structural similarities between AbrBN and AbhN [28, 35]. This study extends previous work to include a third 'AbrB-like' transition-state regulator paralog, SpoVT. SpoVT, from *B. subtilis*, is a transcription factor responsible for the stimulation and/or repression of forespore-specific σ^G-dependent transcription. The forespore is the cellular chamber destined to become the metabolically dormant endospore resulting from the alternative developmental pathway known as sporulation. SpoVT is absolutely required for proper spore formation and viability, as deletion of *spoVT* results in spores exhibiting defective spore coats and reduced resistance properties [39, 40, 43, 66].

Here we present the high resolution residual dipolar coupling (RDC) refined NMR solution structures and dynamic analysis of three of the N-terminal DNA-binding domains of the AbrB family of transcription factors in *B. subtilis*: AbrBN, AbhN and SpoVTN (residues 1-

55). The solution structure of SpoVTN is the first reported structure of SpoVTN and represents the third such structure of the DNA-binding domain of this family of transcription factors. Coupled with residual dipolar coupling (RDC) refinement (which allows for the calculation of more accurate structures, constraining local geometry by restraining bond orientations relative to a common frame) of the previously released structures and dynamic characterization of all three proteins, we are presented with a unique opportunity to analyze, in greater detail than previously available, the structural aspects underpinning the nature of DNA recognition and specificity of this class of proteins. This study also builds upon all known biochemical information (gel shift assays, DNase I footprinting, DNA methylation, and mutational data) with additional AbrB knockout mutational data and chemical shift perturbation experiments to determine a structural model of the AbrB:DNA interaction using the molecular docking program HADDOCK (High Ambiguity Driven protein-protein Docking) [7, 9, 12, 26-28, 32, 34, 81]. This is a significant advance as it is the first detailed structural model for any 'AbrB-like' transition-state regulator binding to a cognate DNA promoter sequence. The model also strongly supports the importance of millisecond timescale motions in the function of 'AbrB-like' transition-state regulators.

Methods

Expression and Purification of AbrBN, AbhN, and SpoVTN

The expression and purification protocols for these three proteins are discussed in detail in Appendix D.

NMR Spectroscopy of SpoVTN

NMR data were collected on a 600 MHz Varian INOVA spectrometer equipped with a triple-resonance probe. Sequential and side chain assignments for SpoVTN were accomplished using the standard suite of assignment experiments described in Appendix A, carried out on 1- 2 mM samples at 305K in 90:10% or 1:99% H_2O:D_2O NMR buffer [82-87]. Hydrogen bond and backbone dihedral angle restraints were obtained as described in Appendix A [88]. NOEs were obtained from three dimensional ^{15}N NOESY-HSQC experiments with 120 and 150ms mixing times as well as ^{13}C NOESY-HSQC and folded ^{13}C NOESY-HSQC experiments with 120ms mixing times [87]. The ^{13}C NOESY-HSQC experiments were carried out on a 600 MHz Varian INOVA equipped with a Varian cryogenic probe. All data was processed and analyzed using NMRPipe and NMR-View on a Linux workstation running Red Hat 9.0 [89, 90].

Residual Dipolar Couplings (RDCs)

RDCs were measured on 1.5 – 3 mM uniformly labeled ^{15}N samples of AbrBN, AbhN, and SpoVTN at 305K in the appropriate 90:10% H_2O:D_2O NMR buffers. HSQC and IPAP-HSQC experiments were recorded on a 500 MHz Varian INOVA equipped with a Varian cryogenic probe at 305 K. 6.0 mm to 4.2 mm (AbrBN) or 5.0 mm to 4.2 mm (AbhN and SpoVTN) radially compressed 7% polyacrylamide gels were used as the alignment media [91]. 1 mL gels were cast in the gel chamber and allowed to fully polymerize. Unpolymerized reagents were removed by dialyzing the gels two times against 50 mL of deionized water. Gels were then cut to approximately 500 μL volume and dehydrated for

two days in a 32°C incubator. Fully dehydrated gels were placed upright into the gel chamber and rehydrated in 500 μL of protein sample. The gel press assembly was then used to insert the 6.0 mm or 5.0 mm gel into an open-ended 4.2 mm NMR tube.

Structure Calculations

For the AbrBN and AbhN structures, previously calculated NMR restraints were used in the structural refinement utilizing the RDC data that was collected [28, 35]. For all three proteins, structure calculations were carried out with NOEs, hydrogen bond restraints, TALOS-predicted φ and ψ angles, and RDCs. The programs ARIA (version 1.2) and CNS (version 1.1) were used to calculate the solution structure starting from an extended polypeptide structure [92, 93]. The CNS protocols used simulated annealing with a combination of torsion angle dynamics and Cartesian dynamics with default parameters. The lowest 10 energy structures were further refined in an explicit water solvent by ARIA. Manually assigned inter- and intramolecular NOEs were input to ARIA as assigned and uncalibrated with respect to distance. The total number of ambiguous NOE restraints allowed for each peak in the NOESY spectra was set to 20. Distance restraints, derived from the manually assigned NOEs, were set to 1.8 – 6.0 Å. The dihedral angle restraints were taken to be 2 S.D. values or at least 20° from the average values predicted by TALOS. In this study, the dihedral angles were restrained to $\varphi = -70 \pm 50°$ and $\psi = -50 \pm 50°$ for the helical regions. The initial and final susceptibility anisotropy (SANI) force constants for the RDC restraints were 0.2 and 1.0 kcal Hz–2, respectively. The final SANI force was adjusted to give average r.m.s. deviation values within a reasonable range of error (approximately 2 Hz).

Da and R values (axial and rhombic components of the molecular alignment tensor, respectively), based on RDC data, within the ARIA protocol were as follows: –13.241 and 0.289 for AbrBN, -4.781 and 0.213 for AbhN and –12.130 and 0.197 for SpoVTN. The non-crystallographic symmetry energy term was used to keep the Cα atoms of the monomers superimposable, and distance symmetry potential was used to ensure that the relative orientations of all of the Cα atoms of the monomers were symmetric. Dipolar coupling restraints for the dimer were fit and analyzed (Q-factor, R-factor and r.m.s. deviation) using singular value decomposition (SVD) with the RDCA program (provided by L.E. Kay, University of Toronto). When carrying out the SVD fit to each the dimeric structures, the measured RDC value was assigned to the N – H vectors of a given residue in each of the two domains.

^{15}N backbone relaxation measurements and model-free analysis of backbone motions

Backbone ^{15}N R_1, R_2 and All-^{15}N NOE relaxation measurements were acquired on a Varian INOVA spectrometer at 9.4 T (399.8636182 MHz) and 305.15 K on uniformly ^{15}N-labeled samples of AbhN, AbrBN and SpoVTN (1-2 mM) with one exception – TCEP (1 mM) was used instead of DTT in the AbhN NMR buffer. R_1 and R_2 measurements were collected with fully interleaved planes acquired at nine and eight relaxation delay values, respectively, with one exception. AbhN T_1 time point measurements were collected individually. Relaxation delays were as follows: AbhN [T_1 = 10 (x2), 50, 100, 160, 220 (x2), 300, 410, 560 (x2), 820 ms], [T_2 = 4 (x2), 18, 34, 54, 76 (x2), 104, 140 (x2), 192 ms]; AbrBN [T_1 = 1.5 (x2), 60, 120, 190, 280 (x2), 380, 520, 710 (x2), 1040 ms], [T_2 = 4 (x2), 16, 32, 54, 80 (x2), 112, 160 (x2),

240 ms]; and SpoVTN [T_1 = 1.5 (x2), 60, 120, 190, 280 (x2), 380, 520, 710 (x2), 1040 ms], [T_2 = 4 (x2), 16, 32, 54, 80 (x2), 112, 160 (x2), 240 ms]. Three replicate $\{^1H\}$-^{15}N NOE spectra were recorded with proton saturation using a 4 s period of saturation and interleaved with the reference spectrum, which was recorded with a 5 s recycle delay and no saturation. The instrument temperature was calibrated using 100% methanol before each set of experiments. Pulse sequences were written in-house from methods previously described [94, 95].

Relaxation data were processed and analyzed using the NMRPipe/NMRDraw software suite using a Lorentz-to-Gauss window function [89, 90]. Peak intensities were quantitated using the nlinLS module of the NMRPipe. R_1 and R_2 relaxation rates were obtained by fitting the peak intensities as a function of relaxation delay time to a single exponential decay function given by

$$I(\tau) = \exp(-R_2\tau)$$

using the Levenberg-Marquardt nonlinear least-squares fitting program Curvefit v1.4 (A. G. Palmer III, Columbia University). Initial values for parameters to be optimized were chosen using grid search methods. Monte Carlo simulations were performed to estimate the uncertainty in the fitted parameters. Experimental errors associated with R_1 and R_2 relaxation rates were estimated from the baseline noise and the calculated uncertainty in peak intensities for duplicate data sets collected with the same relaxation delays [95]. Steady-state $\{^1H\}$-^{15}N

NOE values were calculated as the ratio between cross-peak intensities with (I_{ref}) and without (I_{sat}) ^{1}H saturation (I_{sat}/I_{ref}) averaged over three replicate experiments and errors were obtained by

$$\sigma_{NOE} = [((\sigma_{sat} * I_{ref})^2 + (\sigma_{ref} * I_{sat})^2)/(I_{ref})]^{1/2}$$

For all analyses, ^{15}N relaxation data was applied to both monomers within the dimer structures. The program THESEUS [96] was used to determine the structure most similar to the average structure in the RDC-refined AbhN, AbrBN and SpoVTN NMR structure ensembles. The structures most similar to the average structure were subsequently used for relaxation and model-free analysis. The program PDBINTERTIA (A. G. Palmer III, Columbia University) was used to obtain inertia tensors and structures files rotated so the moments of inertia are aligned with the Cartesian axes and translated so that the center of mass is located at the origin. Diffusion tensor estimates for spherical, axially-symmetric and anisotropic motional models were determined using the programs R2R1_TM and QUADRIC_DIFFUSION (A. G. Palmer III, Columbia University), the later of which was used to produce structures rotated into the diffusion frame for model-free analysis. Residue-specific 10% or 15% trimmed R_2/R_1 ratios were used for the diffusion tensor estimate calculation to account for residues with very fast internal motions ($\tau_e << \tau_m$; low R_2/R_1 ratios or NOE values <0.6) and no significant R_{ex} contribution to R_2 (elevated R_2/R_1 ratios) [97] were also excluded from the diffusion tensor estimate calculation. χ^2 and F statistics reported by QUADRIC_DIFFUSION were initially used to determine the appropriate

diffusion tensor. The program relax [98] was used to fit the experimental ^{15}N relaxation data to five dynamic motional models of the model-free formalism [99-101]: (1) S^2, (2) S^2 and τ_e, (3) S^2 and R_{ex}, (4) S^2, τ_e and R_{ex}, and (5) S_f^2, S^2 and τ_e. The S^2 generalized order parameter is related to the equilibrium distribution of the N-H bond vector orientations; a value of 0 assumes completely unrestricted motion, where as a value of 1 assumes completely restricted motion. The R_{ex} parameter describes the chemical exchange broadening contribution to the R_2 rate constant on the μs-ms time scale. Other parameters, as well as parametric restrictions resulting in the five mathematical motional models described above, have been described previously [97, 102, 103]. Values for the ^{15}N CSA and N-H bond length (r_{NH}) were set to -172 ppm and 1.02 Å, respectively, and Newton minimization were used for all calculations. The quality of the fits between the experimental data and each of the motional models and subsequent model selection was determined using previously described methods of model elimination [104], χ^2 and Akaike's Information Criteria (AIC) statistics [105]:

$$AIC = \chi^2 + 2k$$

where k is the number of model-free parameters in the model and χ^2 describes fit of the relaxation data to the model. A dimer symmetry-forced model selection protocol was used, where the number of datasets (n), k and χ^2 for symmetry equivalent residues were summed before AIC model selection, resulting in the selection of identical models for symmetric residues between the individual monomers of the dimer structure. Final selection of the

appropriate diffusion tensor was determined using the χ^2 and AIC statistics describing the fit of the tensor after convergence to the optimal diffusion tensor parameters, as well as agreement between experimental and back-calculated ^{15}N relaxation parameters. Parameter fitting errors were determined using Monte Carlo simulations (n=500).

Chemical Shift Perturbation - NMR spectroscopy.

^{15}N labeled AbrBN^{53} (residues 1-53) and AbrBN^{55}-tetramer (residues 1-55 form a tetramer upon oxidation of Cys 54) samples used for NMR data collection contained 1 mM protein (~99% pure) and were dialyzed into NMR buffer (Appendix D), DTT being omitted from the NMR buffer for AbrBN^{55}-tetramer. NMR chemical shift perturbations experiments contained ~0.5 mM AbrBN^{55}-tetramer. *abrB8* was titrated into samples in increments that provided the following DNA:protein molar ratios: 0:1, 1:1, 2:1, 4:1, 8:1, and 16:1. ^{1}H-^{15}N HSQC experiments were run at 305 K on a Varian INOVA 600 spectrometer equipped with 4 radiofrequency channels and a single axis pulsed field gradient triple resonance probe. Spectral widths: ^{1}H, 8000 Hz; ^{15}N, 1900 Hz. The carrier was placed on the water resonance in the ^{1}H dimension and at 118.5 ppm in the ^{15}N dimension. 2048 x 128 complex points were recorded in the ^{1}H dimension and ^{15}N dimensions, respectively. Data were referenced to previously published chemical shifts for AbrBN^{53}, processed with NMRPipe and analyzed with NMR-View [89, 90].

HADDOCK Docking Procedure.

The 1Z0R coordinate files for AbrBN[53] were obtained from the RCSB Protein Data Bank (PDB). The coordinate file for AbrBN[55]-tetramer was obtained by modeling the first 55 residues of AbrB against the solved structure of AbrBN[53] by MODELLER (7v2) [106]. The resulting models were analyzed by their respective outputs from PROCHECK and their energy function output from MODELLER [106, 107].

The *abrB8* sequence used in this study is:

5'- ATG ATT GAC AAT TAT TGG AAA CCT -3'

A model of canonical C2'-endo conformation B-DNA were generated by 'DNA Tools' (http://hydra.icgeb.trieste.it/~kristian/dna/) through a trinucleotide base pair parameter with AMBER minimization. Hydrogens were added according to the standard assigning protocol within CNS, followed by a short energy minimization step during the initiation stage in HADDOCK [81, 92].

HADDOCK is comprised of three simulated annealing protocols: (i) rigid docking; (ii) semi-flexible docking and (iii) water refinement. Default HADDOCK parameters were used throughout. A total of 1000 structures were generated in each iteration, with 200 structures being water refined. Each docking attempt was performed 10 times and the solution with the lowest HADDOCK score was retained. Systematic sampling of 180° rotated solutions was used in the rigid-body docking stage to minimize the occurrence of false positives. The

HADDOCK score was defined to rank the structures after each docking stage. It is a weighted sum of intermolecular electrostatic (Elec), vdW, AIR energies, and a buried surface area (BSA) term: rigid-body score = 1.0 * Elec + 1.0 * vdW – 0.05 * BSA + 1.0 * AIR, final score = 1.0 * Elec + 1.0 * vdW + 1.0 * AIR. A cluster analysis was performed on the final docking solutions using a minimum cluster size of 4. The cut-off for clustering was manually determined for each docking run. The root mean square deviation (r.m.s.d.) matrix was calculated over the backbone atoms of the interface residues of the DNA after fitting on the interface residues of the protein.

HADDOCK modeling restraints.

All active/passive residues have a relative solvent accessibility >40% as calculated with NACCESS [108]. Active and passive residues (Table 2.1) were assigned from mutational studies on full length AbrB and chemical shift perturbation experiments. Mutations in the N-terminal domain that disrupt AbrB binding to DNA are: R8S, K9R, G14A, R15H/S, V17A, I18G, P19A, I20L/S, E21A, L22R, R23S, R24S, L26R, I28G/R, L34H and C54S/Y/W. Of these residues Arg 8, Lys 9, Arg 15, Val 17, Ile 20, Glu 21, Leu 22, Arg 23, Arg 24 and Leu 26 are defined as active. Residues Gly 14, Ile 18, Pro 19, Ile 28 and Leu 34 are not solvent accessible and were not directly used in the docking protocol. Residues 1, 5-7, 12-13, 27, 29, 30-32 and 48 are defined as passive. The promoter region of *abrB8* contains an asymmetric TGNNA-5bp-TGGNA binding motif, which has been shown to be a pseudo consensus sequence recognized by AbrB [17]. These base pairs (base pairs 6-10, 16-20, 4'-8', and 14'-18') were defined as the active region while two base pairs on either side of the TGNNA-5bp-TGGNA motif were defined as passive (base pairs 5, 11, 15, 21, 3', 9', 13', and 19') [16].

Table 2.1: Definition of active and passive residues for HADDOCK modeling restraints

	Active	Passive
AbrBN[55]	R8, K9, R15, V17, I20, E21, R23, R24, L26	M1, G5, I6, V7, E12, L13, G27, A29, E30, K31, D32, Y48
abrB8	T23, G24, A25, C26, A27, T33, G34, G35, A36, A37, T177, T178, T179, C180, C181, A186, T187, T188, G189, T190	T22, A28, T32, A38, G176, A182, A185, C191

In order to preserve the helical conformation of DNA the following restraints were defined: planarity restraints for the purine and pyrimidine rings were introduced, and the sugar pucker was restrained to the C2'-endo conformation. Watson–Crick base pairs were defined and hydrogen bond lengths of the input structure were measured and restricted to ±0.05 Å. In a similar way the dihedral angles of the sugar-phosphate backbone of the input structure were measured and used as restraints (restricted to $\alpha = \alpha_{inp} \pm 10°$, $\beta = \beta_{inp} \pm 40°$, $\gamma = \gamma_{inp} \pm 20°$, $\delta = \delta_{inp} \pm 50°$, $\varepsilon = \varepsilon_{inp} \pm 10°$ and $\zeta = \zeta_{inp} \pm 50°$), similar to other published protein:DNA interactions determined through HADDOCK [109]. Finally, phosphorus to phosphorus distances between base pair i and i+1 were measure and restricted to 6.0 – 7.5 Å, values typical for protein:DNA complexes. In addition, in order to preserve the symmetry of the AbrBN[55]-tetramer, a non-crystallographic symmetry energy term (NCS) and a distance symmetry potential were used to keep the Cα atoms of the monomers superimposable. Furthermore, the H-bonding network of the unbound AbrBN[53] was employed as an additional restraint to maintain a similar dimerization interface between the unbound and

bound structures (restricted to: β-sheet, C-N 3.7 ± 0.50nm and O-HN 1.9 ± 0.25nm; α-helix O-HN 2.1 ± 0.25nm and O-N 2.75 ± 0.50nm). Finally, disulfide restraints were used for the disulfide linked AbrBN^{55}-tetramer - restricted to 2.0±0.25nm.

Analysis of resulting modeled structures.

The r.m.s.d. values of the complexes were calculated using ProFit (A.C.R. Martin, www.bioing.org.uk/software/profit). All heavy atoms of the interface residues of the DNA and the protein were used to calculate the r.m.s. deviation of the total complex. The interface was composed of residues 8-32 from the protein and base pairs 23-37 of the DNA. Residues in the flexible termini of the protein were omitted from the calculation. Values for base pair and base pair step parameters as well as torsion angles for the sugar-phosphate backbone and the sugar pucker were obtained using the program 3DNA [110].

Results

RDC Refinements of AbrBN and AbhN

The RDC refined ensemble structures of AbrBN and AbhN are shown in Figure 2.1A through 2.1D. The restraint statistics for the lowest 10 energy structures, generated by the final iteration of ARIA in an explicit water solvent are shown below in Table 2.3 [93]. The RDC refined structures of AbrBN were determined with 1691 unambiguous restraints: 901 intra-residue, 933 sequential ($|i-j|=1$), 461 medium range ($|i-j|<4$), and 788 long range ($|i-j|>5$). The inclusion of 46 RDC restraints allowed for the determination of an ensemble of structures exhibiting a mean backbone and heavy atom r.m.s. deviations of 0.832 ± 0.340 Å

and 1.290 ± 0.281 Å, respectively. The RDC refined structures of AbhN were determined with 1413 unambiguous restraints: 329 intra-residue, 730 sequential (|*i*-*j*|=1), 478 medium range (|*i*-*j*|<4), and 643 long range (|*i*-*j*|>5). After structural refinement with 46 RDC restraints, the ensemble of lowest energy structures exhibited a mean backbone and heavy atom r.m.s. deviations of 0.742 ± 0.227 Å and 1.030 ± 0.205 Å, respectively. In the RDC refined structures of both AbrBN and AbhN secondary structure elements were identical to those previously reported [28, 35].

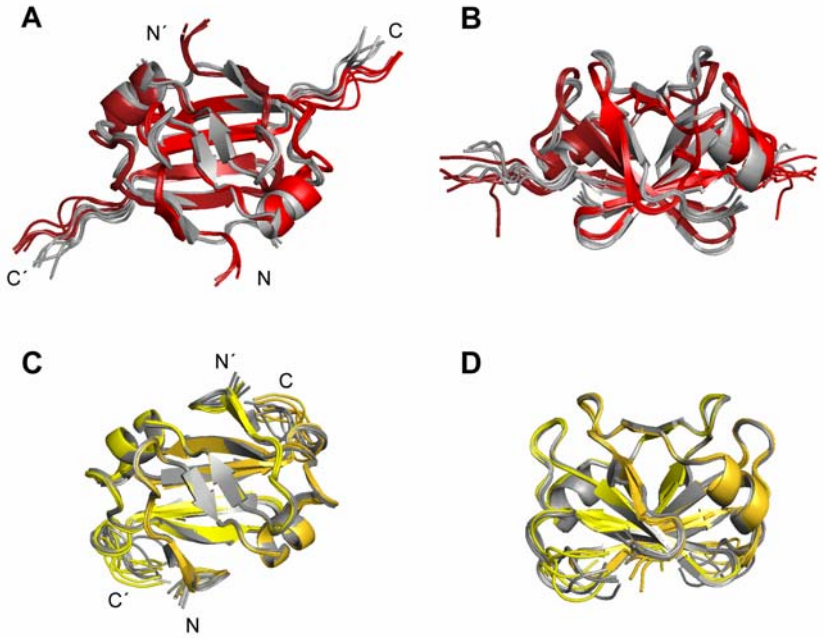


Figure 2.1. Ribbon Diagram of RDC-refined structural ensembles of AbrBN and AbhN. *A* The lowest energy RDC-refined structural ensemble for AbrBN is shown on the top in red, overlaid atop the non-RDC refined structure in gray, *B* The structures in *A* are shown rotated 90^{o} into the plane of the page, *C* The lowest energy RDC-refined structural ensemble for AbhN is shown on the bottom in gold, *D* The structures in *C* are shown rotated 90^{o} into the plane of the page. N- and C-termini are marked for the dimeric subunits.

Structure Determination and RDC refinement of SpoVTN

The $^{1}H - ^{15}N$ HSQC spectrum of SpoVTN (Figure 2.2) showed good peak dispersion with well resolved peaks corresponding to 48 of 55 residues. With the exception of residues 1, 2, and 48, all NH resonances were assigned. All told, 94% of backbone H^{N}, N, C^{α}, and C′ resonances were assigned. The RDC refined structure of SpoVTN (Figure 2.3B) was determined with 2135 unambiguous restraints: 918 intra-residue, 808 sequential ($|i-j|=1$), 621 medium range ($|i-j|<4$), and 935 long range ($|i-j|>5$). In the case of the RDC refined structure of SpoVTN, TALOS/CSI predictions of secondary structure agree well with the energy-minimized structure, being composed of four β-strands (residues 5 – 9, 15 – 17, 34 – 39, and 43 – 47), one α-helix (residues 20 – 26), and three loop regions (loop 1: residues 11 – 14; loop 2: residues 28 – 32; loop 3: residues 40 – 42). These structural elements are illustrated in Figure 2.3A. As described in the previously published structures of AbrBN and AbhN, an extensive β-scaffold exists in SpoVTN – intramolecularly between β-strands 3 and 4 within each monomer, and intermolecularly between β-strands 4 from each monomer. After structural refinement with 38 RDC restraints, the ensemble of lowest energy structures exhibited a mean backbone and heavy atom r.m.s. deviations of 0.603 ± 0.199 Å and 0.986 ± 0.156 Å, respectively. Complete structural statistics for the RDC-refined structures of AbrBN, AbhN, and SpoVTN are shown below in Table 2.2. Chemical shift values for all assigned nuclei are available in Appendix E.

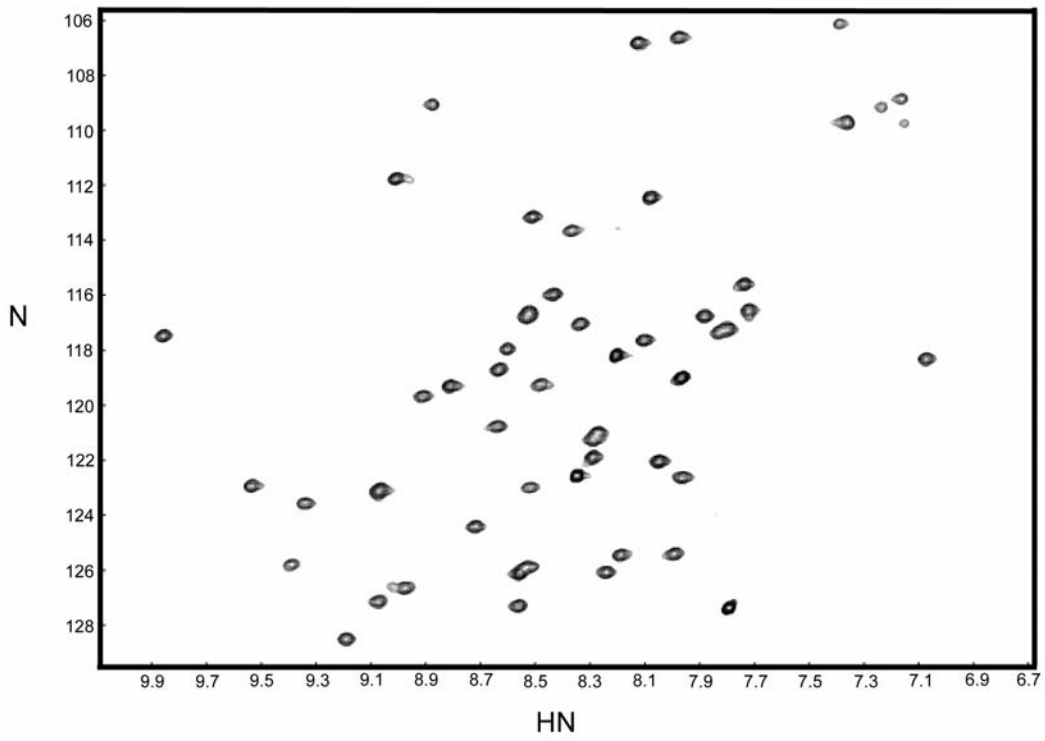


Figure 2.2. ^{1}H-^{15}N HSQC spectrum of 1 mM SpoVTN at 305 K. 10 mM Tris-HCl, 150 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3, pH 5.8.

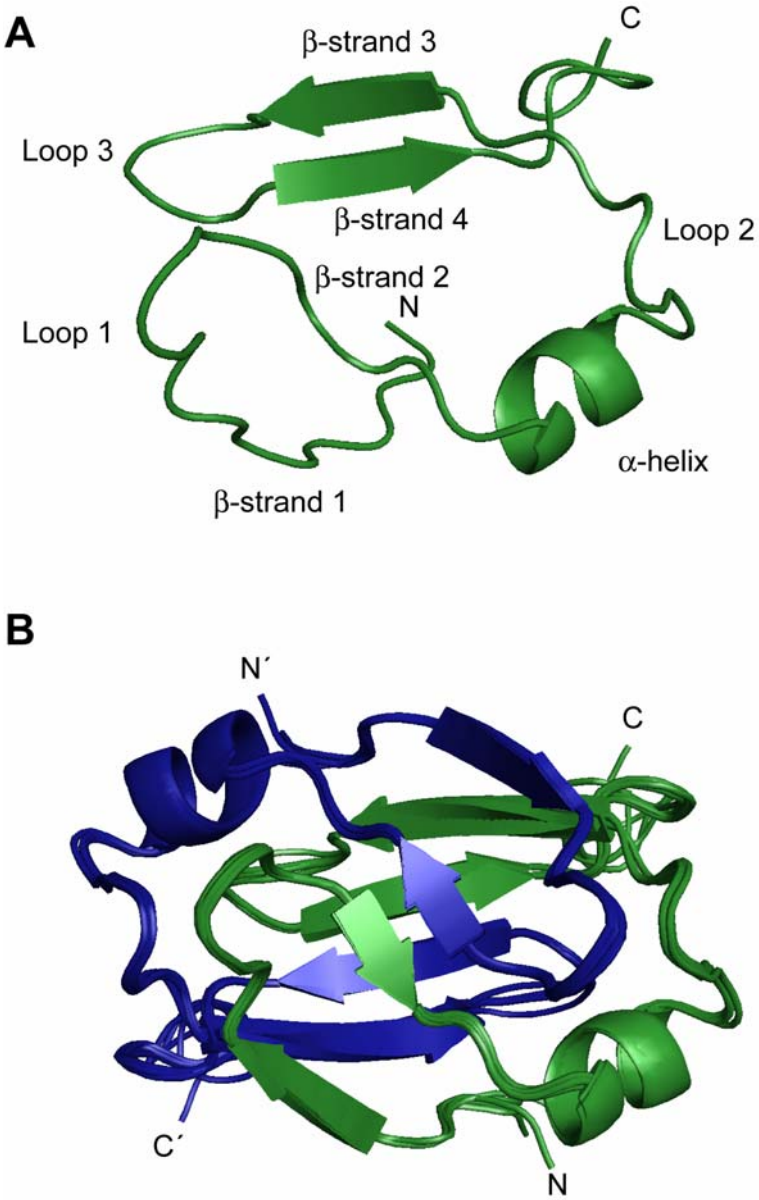


Figure 2.3. RDC-refined structure of SpoVTN. *A* Secondary structure elements are highlighted in the monomeric subunit of SpoVTN. *B* The lowest energy RDC-refined structural ensemble for dimeric SpoVTN. N- and C-termini are marked for each of the monomeric subunits and monomeric subunits are colored blue and green.

Table 2.2. NMR structural refinement statistics for AbrBN, AbhN, and SpoVTN.
a 0.5 Å Violations for the ensemble of 10 lowest energy structures. *b* Pairwise r.m.s. deviation was calculated among the 10 lowest energy structures for residues 1-53 (AbrBN), 1-54 (AbhN) and 1-55 (SpoVTN).

NMR distance and dihedral constraints	AbrBN	AbhN	SpoVTN
Distance constraints			
Total NOE	3083	2180	3282
Ambiguous	1392	767	1147
Unambiguous	1691	1413	2135
Intra-residue	901	329	918
Sequential (|i-j|=1)	933	730	808
Medium range (|i-j|<4)	461	478	621
Long range (|i-j|>5)	788	643	935
Hydrogen bonds	48	70	54
Total dihedral angles			
ϕ	69	47	38
ψ	69	47	38
RDC constraints	46	46	38
Structural statistics			
Energies (kcal mol $^{-1}$)			
van der Waals	-1018.6 ± 32.0	-1016.6 ± 29.8	-1001.7 ± 43.5
Electrostatic	-4044.3 ± 53.0	-4051.1 ± 79.7	-4238.8 ± 79.5
Average violations per structure			
NOEs and/or hydrogen bonds [a]	0	0	0.1
Dihedrals [a]	0	0	0
Violations (mean and S.D)			
Distance constraints (Å)	0.032 ± 0.001	0.030 ± 0.001	0.039 ± 0.001
Hydrogen bonds (Å)	0.055 ± 0.004	0.063 ± 0.004	0.091 ± 0.006
Dihedral angle constraints (°)	0.404 ± 0.134	0.316 ± 0.060	0.783 ± 0.089
Deviations from idealized geometry			
Bond lengths (Å)	0.0042 ± 0.0001	0.0047 ± 0.0001	0.0063 ± 0.0001
Bond angles (°)	0.622 ± 0.014	0.693 ± 0.010	0.813 ± 0.020
Impropers (°)	1.66 ± 0.11	1.91 ± 0.10	1.93 ± 0.07
RDC constraints			
R-factor	0.289	0.213	0.197
Q-factor	0.224	0.207	0.166
R.m.s deviation (Hz)	0.745	0.928	2.217
Correlation (experimental to calculated)	0.999	0.978	0.988
Average pairwise r.m.s (Å) [b]			
Secondary structure (backbone)	0.223 ± 0.049	0.206 ± 0.048	0.221 ± 0.048
Secondary structure (heavy)	0.721 ± 0.096	0.564 ± 0.063	0.632 ± 0.066
Backbone	0.832 ± 0.340	0.742 ± 0.227	0.603 ± 0.199
Heavy	1.29 ± 0.281	1.03 ± 0.205	0.986 ± 0.156
Ramachandran analysis			
Most favored	74.9	71.7	61.5
Additionally allowed	14.1	18.5	26.1
Generously allowed	6.4	7.6	8.7
Disallowed	4.6	2.2	3.7

^{15}N backbone relaxation measurements and model-free analysis of backbone motions

Spin-lattice relaxation (R_1), spin-spin relaxation (R_2) and steady-state $\{^1H\}$-^{15}N nuclear Overhauser effect (NOE) data sets were collected at a magnetic field strength of 9.4 Tesla for 44, 47 and 43 residues of AbhN, AbrBN and SpoVTN out of 54, 53 and 55 total residues, respectively. ^{15}N relaxation data for some residues were not obtained due to unobserved or overlapped resonances, as well as for all proline residues. For AbhN, these residues include Met-1, Lys-2, Val-6, Arg-15, Pro-19, Asp-27, Lys46, Lys-49, Pro-50 and His-51. For AbrBN, these residues include Met-1, Lys-2, Asp-11, Ile-18, Pro-19 and Pro-50. For SpoVTN, these residues include Met-1, Lys-2, Arg-15, Pro-19, Lys-20, Arg-23, Leu-26, Pro-33, Val-38, Ile-45, Lys-48 and Pro-51. The average 10% trimmed relaxation values are as follows: AbhN [R_1 = 2.70(±0.10) s^{-1}, R_2 = 7.87(±0.62) s^{-1}, NOE=0.65(±0.07)]; AbrBN [R_1 = 2.65(±0.10) s^{-1}, R_2 = 7.57(±0.56) s^{-1}, NOE=0.65(±0.06)]; and SpoVTN [R_1 = 2.49(±0.18) s^{-1}, R_2 = 7.74(±1.20) s^{-1}, NOE=0.60(±0.20)].

Residues at the N- and C-terminal regions have relaxation values lower than the average of core residues, indicating considerable motion (very fast internal motions; $\tau_e << \tau_m$) on the picosecond-nanosecond (ps-ns) time scale. Elevated R_2/R_1 or $R_1 \cdot R_2$ values compared to the average core of residues indicate a chemical exchange broadening (R_{ex}) contribution on the μs-ms time scale (i.e., higher frequency motions) [97, 111]. A few residues of AbhN and AbrBN have elevated R_2/R_1 or $R_1 \cdot R_2$ values compared to the average core of residues, including residues Ile-4, Gly-5, Asp-11, Ile-34 and Val-38 for AbhN and residues Tyr-37 and

Asp-41 for AbrBN. For SpoVTN, no R_2/R_1 or $R_1 \cdot R_2$ values are significantly elevated compared to the average core of residues. Measurements made on ½ or ⅓ diluted samples indicated that there were no significant concentration-dependent effects on the ^{15}N relaxation rates.

Initial estimates for isotropic, axially symmetric and anisotropic diffusion tensors were obtained using the ^{15}N relaxation rates measured at 9.4 Tesla. The relative moments of inertia for the AbhN, AbrBN and SpoVTN RDC-refined structures are (1.00:0.96:0.75), (1.00:0.95:0.69) and (1.00:0.94:0.65), respectively, suggesting that axially symmetric tensors may be appropriate diffusion tensors for these proteins. RDC-refined NMR structures were used to calculate initial parameter estimates and fitting statistics for the isotropic, axially symmetric and anisotropic rotational diffusion tensors. For AbhN, the *F* statistics reported were 3.14 (isotropic vs. axial prolate) and 2.05 (axial prolate vs. anisotropic). These compare with tabulated critical values of $F_{0.05}(3,66) = 2.74$ and $F_{0.05}(2,64) = 3.14$, respectively, which suggest that the use of an axial prolate tensor is a slight improvement over an isotropic tensor, whereas the anisotropic tensor is not a statistically significant improvement over the axial prolate tensor. For AbrBN, the *F* statistics reported were 8.84 (isotropic vs. axial prolate) and <0.001 (axial prolate vs. anisotropic). These compare with tabulated critical values of $F_{0.05}(3,56) = 2.77$ and $F_{0.05}(2,54) = 3.17$, respectively, which suggest that the use of an axial prolate tensor is an improvement over an isotropic tensor, whereas the anisotropic tensor is not a statistically significant improvement over the axial prolate tensor. For SpoVTN, the *F* statistics reported were 13.16 (isotropic vs. axial prolate) and <0.001 (axial

prolate vs. anisotropic). These compare with tabulated critical values of $F_{0.05}(3,54) = 2.78$ and $F_{0.05}(2,52) = 3.18$, respectively, which suggest that the use of an axial prolate tensor is an improvement over an isotropic tensor, whereas the anisotropic tensor is not a statistically significant improvement over the axial prolate tensor.

Isotropic or axial prolate tensor parameter estimates were subsequently used for model-free analysis of internal motion parameters. The initial global rotational correlation time (τ_c) estimates and $D_{\parallel}/D_{\perp}$ values were set as follows: AbhN [isotropic: $\tau_c = 6.54$ ns; axial prolate: $\tau_c = 6.57$ ns; $D_{\parallel}/D_{\perp} = 1.18$], AbrBN [isotropic: $\tau_c = 6.39$ ns; axial prolate: $\tau_c = 6.35$ ns; $D_{\parallel}/D_{\perp} = 1.11$], SpoVTN [isotropic: $\tau_c = 6.91$ ns; axial prolate: $\tau_c = 6.94$ ns; $D_{\parallel}/D_{\perp} = 1.22$]. The higher τ_c value for SpoVTN is consistent with the decreased R_1 and increased R_2 values compared to AbhN and AbrBN. The τ_c values are consistent with the molecular size 10-12 kDa, supporting other bioanalytical methods suggesting a dimeric molecular state in solution [25, 35, 43, 44].

Several criteria were used to determine the diffusion tensor model and parameters most appropriate for fitting the relaxation data, including (1) *F* statistics describing the tensor estimations, as detailed above; (2) χ^2 and AIC statistics describing the fit of the relaxation data when assuming an isotropic or axial prolate tensor; and (3) the agreement between experimental and independently back-calculated ^{15}N relaxation parameters. Based on the agreement among these criteria, the isotropic tensor determined to be most appropriate for

AbhN (τ_c = 6.38 ns), whereas an axial prolate tensor was most appropriate for AbrBN (τ_c = 6.36 ns; $D_{\parallel}/D_{\perp}$ = 1.14; ϕ= 172.10°; φ = 72.06°) and SpoVTN (τ_c = 7.01 ns; $D_{\parallel}/D_{\perp}$ = 1.21; ϕ= 170.73°; φ = 2.14°). Figure 2.4 below shows the model-free dynamic parameters S^2 and R_{ex} for all three proteins. S^2 varies between 0 and 1, with a value of 1 represents completely restricted motion. Complete dynamic parameters for all three proteins are available below in Tables 2.3 – 2.5.

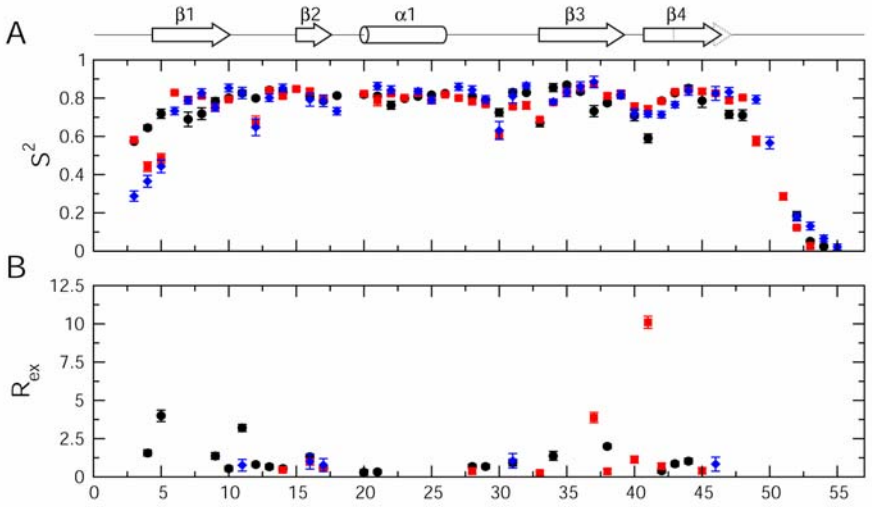


Figure 2.4. Model-free dynamic parameters. *A* S^2 and *B* R_{ex} dynamic parameters plotted vs. residue number for AbhN (black circles), AbrBN (red squares) and SpoVTN (blue diamonds). The positions of the α-helical and β-strand secondary structure are indicated schematically at the top. Error bars indicate the fitting errors.

Table 2.3. AbrBN ^{15}N relaxation and dynamics parameters (pH 5.8, 305.15 K, 399.863182 MHz)

Residue	R_1 (s^{-1})	R_2 (s^{-1})	NOE	Model	S^2	S_f^2	$\tau_e < 100$ or τ_f (ps)	$\tau_e > 100$ or τ_s (ps)	R_{ex} (s^{-1})
M1									
K2									
S3	2.1736 ± 0.0639	5.5569 ± 0.1836	0.2904 ± 0.0293	$m2$	0.582 ± 0.013			102.60 ± 9.15	
T4	2.3490 ± 0.0345	5.2894 ± 0.1210	0.4162 ± 0.0169	$m5$	0.442 ± 0.025	0.759 ± 0.010		2059.86 ± 177.25	
G5	2.5043 ± 0.0280	5.7513 ± 0.1139	0.5022 ± 0.0142	$m5$	0.484 ± 0.026	0.786 ± 0.009		2496.21 ± 226.76	
I6	2.6861 ± 0.0261	8.0891 ± 0.1156	0.7184 ± 0.0123	$m2$	0.829 ± 0.007		19.90 ± 7.70		
V7	2.6706 ± 0.0325	7.4461 ± 0.1193	0.6637 ± 0.0173	$m2$	0.793 ± 0.008		42.04 ± 9.57		
R8	2.7655 ± 0.0357	7.4413 ± 0.1340	0.6742 ± 0.0180	$m2$	0.812 ± 0.008		43.22 ± 10.73		
K9	2.5609 ± 0.0370	7.1692 ± 0.1457	0.6148 ± 0.0233	$m2$	0.755 ± 0.009		55.93 ± 10.64		
V10	2.5997 ± 0.0298	7.7107 ± 0.1348	0.6856 ± 0.0175	$m2$	0.793 ± 0.008		33.34 ± 8.72		
D11									
E12	2.6574 ± 0.0332	7.1073 ± 0.1115	0.6807 ± 0.0128	$m5$	0.675 ± 0.031	0.811 ± 0.009		3694.34 ± 1049.21	
L13	2.8123 ± 0.0326	7.7957 ± 0.1169	0.7160 ± 0.0154	$m2$	0.841 ± 0.008		22.56 ± 10.80		
G14	2.7254 ± 0.0369	7.9851 ± 0.1242	0.6912 ± 0.0161	$m4$	0.811 ± 0.011		30.35 ± 9.30		0.457 ± 0.169
R15	2.8428 ± 0.0373	7.8769 ± 0.1264	0.7007 ± 0.0152	$m2$	0.848 ± 0.008		32.03 ± 12.09		
V16	2.7288 ± 0.0399	8.9475 ± 0.1741	0.7695 ± 0.0246	$m3$	0.836 ± 0.012				0.982 ± 0.213
V17	2.6109 ± 0.0320	8.3525 ± 0.1448	0.6745 ± 0.0182	$m4$	0.797 ± 0.010		40.95 ± 10.07		0.571 ± 0.168
I18									
P19									
I20	2.6920 ± 0.0540	7.9512 ± 0.1775	0.7054 ± 0.0212	$m2$	0.823 ± 0.011		27.53 ± 13.65		
E21	2.7764 ± 0.0363	7.9375 ± 0.1300	0.6762 ± 0.0178	$m5$	0.787 ± 0.027	0.872 ± 0.021		2044.46 ± 891.82	
L22	2.7735 ± 0.0333	7.6967 ± 0.1260	0.6757 ± 0.0182	$m2$	0.824 ± 0.009		47.71 ± 12.62		
R23	2.6596 ± 0.0357	7.7223 ± 0.1409	0.6624 ± 0.0189	$m2$	0.802 ± 0.009		48.23 ± 10.15		
R24	2.6914 ± 0.0298	8.0801 ± 0.1266	0.6892 ± 0.0149	$m2$	0.826 ± 0.007		39.21 ± 8.80		
T25	2.5899 ± 0.0266	7.5994 ± 0.1109	0.6842 ± 0.0133	$m2$	0.787 ± 0.007		33.43 ± 6.63		
L26	2.7249 ± 0.0344	7.7575 ± 0.1290	0.6851 ± 0.0189	$m2$	0.819 ± 0.009		39.75 ± 11.56		
G27	2.6555 ± 0.0291	7.5830 ± 0.1241	0.6996 ± 0.0155	$m2$	0.801 ± 0.008		27.54 ± 8.52		
I28	2.5833 ± 0.0300	7.8689 ± 0.1260	0.6843 ± 0.0166	$m4$	0.782 ± 0.010		31.22 ± 8.12		0.374 ± 0.153
A29	2.5980 ± 0.0324	7.1984 ± 0.1239	0.6488 ± 0.0161	$m2$	0.768 ± 0.008		45.86 ± 7.21		
E30	2.5871 ± 0.0254	6.6108 ± 0.0979	0.5891 ± 0.0102	$m5$	0.606 ± 0.021	0.810 ± 0.007		2481.58 ± 277.52	
K31	2.6082 ± 0.0425	7.1136 ± 0.1420	0.5713 ± 0.0229	$m2$	0.756 ± 0.009		78.61 ± 10.89		
D32	2.7195 ± 0.0260	7.5300 ± 0.1001	0.6602 ± 0.0102	$m5$	0.762 ± 0.020	0.848 ± 0.018		1808.72 ± 531.15	
A33	2.3553 ± 0.0249	6.8766 ± 0.1019	0.5536 ± 0.0115	$m4$	0.688 ± 0.007		61.72 ± 3.86		0.251 ± 0.118
L34	2.6205 ± 0.0335	7.2779 ± 0.1374	0.6681 ± 0.0186	$m2$	0.778 ± 0.008		36.73 ± 8.92		
E35	2.7932 ± 0.0354	7.8011 ± 0.1404	0.6795 ± 0.0191	$m2$	0.832 ± 0.009		42.77 ± 13.60		
I36	2.8082 ± 0.0442	8.1063 ± 0.1730	0.7093 ± 0.0263	$m1$	0.856 ± 0.010				
Y37	2.8901 ± 0.0594	12.0044 ± 0.2950	0.6978 ± 0.0366	$m3$	0.873 ± 0.018				3.878 ± 0.335
V38	2.6849 ± 0.0356	8.0382 ± 0.1367	0.7107 ± 0.0207	$m4$	0.812 ± 0.011		20.69 ± 12.25		0.358 ± 0.167
D39	2.7464 ± 0.0308	7.8259 ± 0.1306	0.6843 ± 0.0173	$m2$	0.825 ± 0.008		38.95 ± 11.27		
D40	2.5102 ± 0.0486	8.4116 ± 0.1781	0.6758 ± 0.0212	$m4$	0.758 ± 0.015		31.14 ± 9.34		1.130 ± 0.229
E41	2.5470 ± 0.0497	17.0647 ± 0.4005	0.6112 ± 0.0279	$m4$	0.744 ± 0.016		54.93 ± 11.82		10.102 ± 0.397
K42	2.6670 ± 0.0406	8.1077 ± 0.1609	0.6159 ± 0.0271	$m4$	0.785 ± 0.013		67.42 ± 14.53		0.697 ± 0.199
I43	2.7464 ± 0.0363	8.0444 ± 0.1535	0.6827 ± 0.0223	$m2$	0.833 ± 0.010		40.67 ± 15.47		
I44	2.8114 ± 0.0394	7.9083 ± 0.1546	0.6959 ± 0.0226	$m2$	0.843 ± 0.010		34.34 ± 16.63		
L45	2.7702 ± 0.0422	8.2860 ± 0.1588	0.7015 ± 0.0249	$m4$	0.835 ± 0.013		30.57 ± 16.97		0.398 ± 0.200
K46	2.7446 ± 0.0467	7.9258 ± 0.1749	0.6608 ± 0.0305	$m2$	0.826 ± 0.012		50.92 ± 20.29		
K47	2.6524 ± 0.0373	7.3669 ± 0.1394	0.6653 ± 0.0227	$m2$	0.787 ± 0.010		39.52 ± 12.21		
Y48	2.7187 ± 0.0348	7.5128 ± 0.1194	0.6470 ± 0.0171	$m2$	0.803 ± 0.008		55.41 ± 10.20		
K49	2.3590 ± 0.0321	5.8393 ± 0.1224	0.3802 ± 0.0184	$m5$	0.576 ± 0.025	0.781 ± 0.023		1065.52 ± 245.75	
P50									
N51	1.7461 ± 0.0269	3.6166 ± 0.1041	-0.2519 ± 0.0160	$m5$	0.286 ± 0.019	0.677 ± 0.010		946.92 ± 82.50	
M52	1.2960 ± 0.0191	2.1434 ± 0.0844	-2.1786 ± 0.0478	$m2$	0.123 ± 0.005			225.42 ± 6.87	
T53	0.7918 ± 0.0102	0.9885 ± 0.0697	-2.3060 ± 0.0144	$m5$	0.027 ± 0.012	0.466 ± 0.015		493.22 ± 61.13	

Table 2.4. AbhN ^{15}N relaxation and dynamics parameters (pH 5.5, 305.15 K, 399.863182 MHz)

Residue	R_1 (s^{-1})	R_2 (s^{-1})	NOE	Model	S^2	S_f^2	$\tau_e < 100$ or τ_f (ps)	$\tau_e > 100$ or τ_s (ps)	R_{ex} (s^{-1})
M1									
K2									
S3	2.1509 ± 0.0386	5.6579 ± 0.1012	0.2156 ± 0.0125	$m2$	0.574 ± 0.007			119.56 ± 5.19	
I4	2.3507 ± 0.0507	7.8683 ± 0.1497	0.3230 ± 0.0230	$m4$	0.646 ± 0.013			126.31 ± 10.96	1.565 ± 0.198
G5	2.5445 ± 0.0767	10.9287 ± 0.2998	0.4296 ± 0.0365	$m4$	0.718 ± 0.024			128.70 ± 23.01	3.991 ± 0.374
V6									
V7	2.7089 ± 0.0476	7.0747 ± 0.1310	0.6761 ± 0.0247	$m5$	0.689 ± 0.038	0.820 ± 0.015		3496.73 ± 1592.72	
R8	2.8617 ± 0.0392	7.4218 ± 0.1085	0.6662 ± 0.0177	$m5$	0.719 ± 0.031	0.869 ± 0.011		3311.23 ± 963.46	
K9	2.6442 ± 0.0570	8.8156 ± 0.1852	0.6399 ± 0.0311	$m4$	0.785 ± 0.017		54.66 ± 15.48		1.369 ± 0.219
V10	2.6814 ± 0.0360	8.1236 ± 0.0970	0.6682 ± 0.0159	$m4$	0.801 ± 0.011		43.59 ± 9.03		0.541 ± 0.159
D11	2.7421 ± 0.0466	10.9938 ± 0.1796	0.7061 ± 0.0245	$m4$	0.827 ± 0.014		25.76 ± 14.47		3.196 ± 0.232
E12	2.7015 ± 0.0312	8.4019 ± 0.0903	0.6286 ± 0.0123	$m4$	0.800 ± 0.009		67.00 ± 7.66		0.807 ± 0.122
L13	2.8206 ± 0.0382	8.6336 ± 0.1116	0.6670 ± 0.0165	$m4$	0.843 ± 0.011		59.36 ± 12.31		0.659 ± 0.158
G14	2.7583 ± 0.0392	8.3976 ± 0.1121	0.7063 ± 0.0166	$m4$	0.831 ± 0.012		26.54 ± 11.27		0.553 ± 0.146
R15									
I16	2.7119 ± 0.0425	9.0149 ± 0.1292	0.6875 ± 0.0216	$m4$	0.814 ± 0.013		35.16 ± 13.23		1.324 ± 0.170
V17	2.6244 ± 0.0351	8.0345 ± 0.0958	0.6790 ± 0.0164	$m4$	0.786 ± 0.012		33.93 ± 8.19		0.601 ± 0.135
M18	2.6992 ± 0.0363	7.7538 ± 0.0960	0.6711 ± 0.0165	$m2$	0.814 ± 0.008		44.38 ± 9.44		
P19									
I20	2.7647 ± 0.0523	8.0738 ± 0.1215	0.6384 ± 0.0156	$m4$	0.820 ± 0.015		70.18 ± 12.48		0.290 ± 0.176
E21	2.6996 ± 0.0300	7.9948 ± 0.0866	0.6984 ± 0.0118	$m4$	0.812 ± 0.009		28.06 ± 7.28		0.326 ± 0.118
L22	2.8088 ± 0.0371	7.5371 ± 0.0958	0.6448 ± 0.0152	$m5$	0.763 ± 0.021	0.871 ± 0.013		1921.13 ± 553.72	
R23	2.7037 ± 0.0329	7.5052 ± 0.0975	0.6484 ± 0.0149	$m2$	0.798 ± 0.007		55.25 ± 8.51		
R24	2.7084 ± 0.0293	7.7029 ± 0.0890	0.6564 ± 0.0126	$m2$	0.810 ± 0.006		53.16 ± 7.45		
A25	2.6996 ± 0.0279	7.7550 ± 0.0810	0.7016 ± 0.0110	$m2$	0.817 ± 0.005		26.45 ± 6.76		
L26	2.7406 ± 0.0384	7.8232 ± 0.0962	0.6840 ± 0.0176	$m2$	0.825 ± 0.008		39.77 ± 12.20		
D27									
I28	2.6821 ± 0.0394	8.3049 ± 0.1076	0.7090 ± 0.0183	$m4$	0.809 ± 0.013		21.16 ± 10.19		0.674 ± 0.161
A29	2.6426 ± 0.0390	8.1385 ± 0.1081	0.6581 ± 0.0181	$m4$	0.788 ± 0.012		45.55 ± 9.83		0.677 ± 0.145
I30	2.7079 ± 0.0305	7.2239 ± 0.0795	0.6614 ± 0.0117	$m5$	0.725 ± 0.020	0.833 ± 0.008		2395.55 ± 563.33	
K31	2.7742 ± 0.0520	8.7863 ± 0.1528	0.6753 ± 0.0270	$m4$	0.830 ± 0.017		48.08 ± 18.59		0.933 ± 0.223
D32	2.7421 ± 0.0308	7.8682 ± 0.0894	0.6875 ± 0.0119	$m2$	0.828 ± 0.007		38.07 ± 8.16		
S33	2.5340 ± 0.0324	6.7325 ± 0.0927	0.6645 ± 0.0161	$m5$	0.673 ± 0.023	0.777 ± 0.010		2586.83 ± 765.99	
I34	2.8656 ± 0.0659	9.4718 ± 0.1875	0.6636 ± 0.0300	$m4$	0.855 ± 0.021		68.95 ± 28.34		1.374 ± 0.290
E35	2.9311 ± 0.0397	8.1240 ± 0.1090	0.6876 ± 0.0178	$m2$	0.869 ± 0.009		56.66 ± 17.73		
F36	2.8158 ± 0.0475	7.7960 ± 0.1278	0.6874 ± 0.0233	$m2$	0.834 ± 0.011		43.07 ± 15.68		
F37	2.7870 ± 0.0400	7.3773 ± 0.1110	0.6755 ± 0.0188	$m5$	0.732 ± 0.029	0.849 ± 0.015		3042.61 ± 1162.28	
V38	2.5948 ± 0.0362	9.3404 ± 0.1128	0.6687 ± 0.0178	$m4$	0.775 ± 0.011		36.97 ± 8.77		2.002 ± 0.146
D39	2.7327 ± 0.0362	7.6620 ± 0.0952	0.6843 ± 0.0159	$m2$	0.815 ± 0.008		38.11 ± 10.34		
G40	2.5607 ± 0.0467	6.9509 ± 0.1153	0.6772 ± 0.0156	$m5$	0.709 ± 0.027	0.787 ± 0.043		2260.86 ± 1198.03	
D41	2.5195 ± 0.0335	6.3131 ± 0.0844	0.6365 ± 0.0140	$m5$	0.590 ± 0.023	0.766 ± 0.009		3216.20 ± 532.24	
K42	2.6555 ± 0.0363	7.8631 ± 0.1017	0.6350 ± 0.0166	$m4$	0.787 ± 0.011		58.25 ± 9.55		0.391 ± 0.149
I43	2.7765 ± 0.0386	8.6892 ± 0.1164	0.6569 ± 0.0196	$m4$	0.827 ± 0.012		60.23 ± 13.62		0.851 ± 0.161
I44	2.8395 ± 0.0425	9.0868 ± 0.1314	0.6865 ± 0.0220	$m4$	0.852 ± 0.015		47.40 ± 17.11		1.035 ± 0.184
L45	2.8579 ± 0.0511	7.7720 ± 0.1295	0.7111 ± 0.0252	$m5$	0.787 ± 0.035	0.865 ± 0.028		3937.87 ± 1981.98	
K46									
K47	2.7537 ± 0.0360	7.2083 ± 0.0999	0.6075 ± 0.0167	$m5$	0.715 ± 0.021	0.860 ± 0.011		1922.11 ± 378.37	
Y48	2.8577 ± 0.0370	7.3659 ± 0.1015	0.6523 ± 0.0169	$m5$	0.711 ± 0.028	0.871 ± 0.010		3030.85 ± 750.37	
K49									
P50									
H51									
G52	1.4015 ± 0.0448	2.5903 ± 0.1077	-0.7742 ± 0.0180	$m5$	0.185 ± 0.022	0.607 ± 0.018		732.17 ± 100.40	
V53	1.1574 ± 0.0170	1.5452 ± 0.0502	-1.3815 ± 0.0087	$m5$	0.051 ± 0.010	0.558 ± 0.008		807.82 ± 33.11	
C54	0.8344 ± 0.0134	1.0333 ± 0.0448	-1.7870 ± 0.0090	$m5$	0.025 ± 0.009	0.434 ± 0.007		692.23 ± 36.47	

Table 2.5. SpoVTN ^{15}N relaxation and dynamics parameters (pH 5.8, 305.15 K, 399.863182 MHz)

Residue	R_1 (s^{-1})	R_2 (s^{-1})	NOE	Model	S^2	S_f^2	$\tau_e < 100$ or τ_f (ps)	$\tau_e > 100$ or τ_s (ps)	R_{ex} (s^{-1})
M1									
K2									
A3	1.8843 ± 0.0585	3.9227 ± 0.1397	-0.2724 ± 0.0328	$m5$	0.288 ± 0.027	0.739 ± 0.021		1015.14 ± 109.60	
T4	1.9268 ± 0.0661	4.4573 ± 0.1616	0.0420 ± 0.0379	$m5$	0.365 ± 0.031	0.707 ± 0.021		1125.42 ± 180.58	
G5	2.2233 ± 0.0726	5.3698 ± 0.1714	0.2845 ± 0.0360	$m5$	0.443 ± 0.033	0.768 ± 0.023		1500.83 ± 222.14	
I6	2.4278 ± 0.0893	7.3560 ± 0.2311	0.6435 ± 0.0518	$m2$	0.733 ± 0.019		45.36 ± 18.84		
V7	2.5582 ± 0.0909	7.7004 ± 0.2427	0.7259 ± 0.0574	$m1$	0.788 ± 0.019				
R8	2.6888 ± 0.1085	8.0422 ± 0.2992	0.6739 ± 0.0609	$m1$	0.827 ± 0.022				
R9	2.4357 ± 0.0849	7.4805 ± 0.2422	0.6585 ± 0.0478	$m2$	0.751 ± 0.020		38.17 ± 18.87		
I10	2.6296 ± 0.0940	8.6352 ± 0.2500	0.7009 ± 0.0501	$m1$	0.853 ± 0.020				
D11	2.5931 ± 0.0906	9.1125 ± 0.2477	0.7043 ± 0.0487	$m3$	0.827 ± 0.030				0.760 ± 0.380
D12	2.5398 ± 0.0811	7.1164 ± 0.1874	0.6441 ± 0.0405	$m5$	0.648 ± 0.044	0.803 ± 0.027		2751.94 ± 1445.58	
L13	2.6186 ± 0.0943	7.9803 ± 0.2304	0.6720 ± 0.0430	$m2$	0.802 ± 0.019		46.42 ± 23.83		
G14	2.6304 ± 0.0942	8.6823 ± 0.2773	0.6941 ± 0.0515	$m1$	0.853 ± 0.020				
R15									
V16	2.4651 ± 0.1130	9.0874 ± 0.3423	0.6826 ± 0.0712	$m3$	0.792 ± 0.034				0.999 ± 0.491
V17	2.4475 ± 0.0932	9.1050 ± 0.2816	0.6592 ± 0.0581	$m4$	0.786 ± 0.030		49.46 ± 28.32		0.773 ± 0.419
I18	2.4517 ± 0.0946	7.3903 ± 0.2335	0.5342 ± 0.0458	$m2$	0.732 ± 0.019		86.74 ± 21.28		
P19									
K20									
E21	2.6037 ± 0.0926	9.1070 ± 0.2776	0.7578 ± 0.0470	$m1$	0.863 ± 0.019				
I22	2.7066 ± 0.1076	8.3331 ± 0.2683	0.6902 ± 0.0577	$m1$	0.843 ± 0.021				
R23									
R24	2.5287 ± 0.0757	8.7864 ± 0.2197	0.6996 ± 0.0431	$m1$	0.836 ± 0.015				
T25	2.4742 ± 0.0818	8.2997 ± 0.2306	0.6888 ± 0.0440	$m2$	0.790 ± 0.018		37.58 ± 19.88		
L26									
R27	2.7174 ± 0.0938	8.6176 ± 0.2617	0.7074 ± 0.0572	$m1$	0.859 ± 0.019				
I28	2.5454 ± 0.0916	8.8893 ± 0.2865	0.6798 ± 0.0551	$m1$	0.843 ± 0.021				
R29	2.5648 ± 0.0966	7.9096 ± 0.2506	0.6601 ± 0.0545	$m2$	0.793 ± 0.020		47.45 ± 27.45		
E30	2.5428 ± 0.0665	7.2642 ± 0.1820	0.7038 ± 0.0346	$m5$	0.631 ± 0.047	0.787 ± 0.017		5182.31 ± 1827.53	
G31	2.5793 ± 0.1098	9.0480 ± 0.3376	0.7597 ± 0.0742	$m3$	0.810 ± 0.035				1.051 ± 0.473
D32	2.6584 ± 0.0771	8.8369 ± 0.1931	0.7491 ± 0.0375	$m1$	0.866 ± 0.014				
P33									
L34	2.5446 ± 0.0672	7.8185 ± 0.1675	0.6235 ± 0.0268	$m2$	0.781 ± 0.013		63.42 ± 14.58		
E35	2.6724 ± 0.1014	8.3764 ± 0.2871	0.6459 ± 0.0565	$m2$	0.831 ± 0.023		69.99 ± 43.10		
I36	2.7277 ± 0.1324	8.5371 ± 0.3782	0.7002 ± 0.0817	$m1$	0.860 ± 0.026				
F37	2.7296 ± 0.1279	9.0350 ± 0.4125	0.6993 ± 0.0878	$m1$	0.885 ± 0.029				
V38									
D39	2.6028 ± 0.1000	8.2973 ± 0.2807	0.6468 ± 0.0634	$m2$	0.818 ± 0.021		59.80 ± 38.52		
R40	2.4573 ± 0.1157	7.0100 ± 0.3032	0.6030 ± 0.0586	$m2$	0.720 ± 0.024		56.25 ± 20.32		
D41	2.3899 ± 0.0756	7.1685 ± 0.2100	0.5949 ± 0.0446	$m2$	0.717 ± 0.016		57.84 ± 15.70		
G42	2.3384 ± 0.0785	7.1420 ± 0.2137	0.6148 ± 0.0468	$m2$	0.714 ± 0.017		47.15 ± 16.51		
E43	2.5635 ± 0.0790	7.5426 ± 0.2070	0.6275 ± 0.0410	$m2$	0.766 ± 0.016		59.59 ± 19.69		
V44	2.5749 ± 0.1037	8.5439 ± 0.2982	0.6880 ± 0.0614	$m1$	0.838 ± 0.023				
I45									
L46	2.6197 ± 0.1139	9.0040 ± 0.3239	0.7291 ± 0.0729	$m3$	0.825 ± 0.036				0.834 ± 0.458
K47	2.6247 ± 0.1105	8.2852 ± 0.2930	0.6917 ± 0.0669	$m1$	0.832 ± 0.024				
K48									
Y49	2.5876 ± 0.0975	7.9419 ± 0.2593	0.6812 ± 0.0606	$m2$	0.793 ± 0.021		42.84 ± 29.28		
S50	2.2354 ± 0.0667	6.1094 ± 0.1645	0.3980 ± 0.0310	$m5$	0.566 ± 0.031	0.763 ± 0.053		1108.49 ± 387.02	
P51									
I52	1.7536 ± 0.0428	3.1045 ± 0.1072	-0.3885 ± 0.0194	$m5$	0.178 ± 0.020	0.697 ± 0.016		1186.61 ± 66.48	
S53	1.5718 ± 0.0466	2.5583 ± 0.1081	-0.8862 ± 0.0301	$m5$	0.131 ± 0.020	0.694 ± 0.020		906.25 ± 63.48	
E54	1.3012 ± 0.0398	1.8017 ± 0.0992	-1.7582 ± 0.0475	$m5$	0.066 ± 0.017	0.681 ± 0.022		629.44 ± 55.55	
L55	0.9021 ± 0.0340	1.1177 ± 0.0912	-1.6549 ± 0.0244	$m5$	0.022 ± 0.015	0.458 ± 0.017		755.14 ± 55.14	

DNA bound model.

Previous studies have shown that: (i) full length AbrB functions as a tetramer; (ii) dimeric AbrBN^{55} can be oxidized to form a tetramer (AbrBN^{55}-tetramer via a Cys54-Cys54′ disulfide bond); and (iii) that the AbrBN^{55}-tetramer binds to cognate promoters of full length AbrB with nearly wild-type affinity [20, 22, 26, 27]. This suggests that the binding mechanism of the AbrBN^{55}-tetramer is analogous to that of full length AbrB. To confirm that no structural differences exist between dimeric AbrBN^{53} and AbrBN^{55}-tetramer (apart from the multimerization), ^{1}H-^{15}N HSQC NMR spectra of both are shown below in Figure 2.5. The spectra of AbrBN^{53} and AbrBN^{55}-tetramer overlay extremely well, indicating that disulfide bond formation and tetramerization do not affect the structural unit of AbrBN^{55}-tetramer. Therefore, a structural model of the AbrBN^{55}-tetramer was constructed from the previously solved high-resolution AbrBN^{53} structure [28].

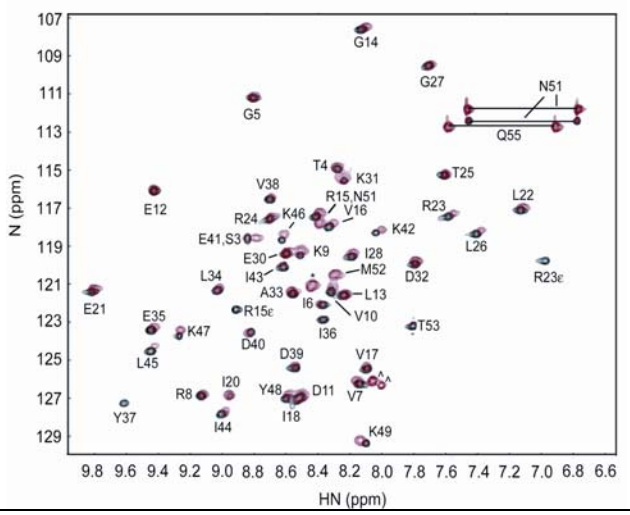


Figure 2.5. ^{1}H-^{15}N HSQC spectra overlay of AbrBN^{53} and oxidized AbrBN^{55}. AbrBN^{53} spectra denoted with black peaks while AbrBN^{55} spectra denoted with red peaks. Spectrum is labeled with AbrBN^{53} assignments. Gln55 is only present in AbrBN^{55} as it is the 55th residue. "^" denotes extra peaks in the AbrBN^{55} spectra that can be attributed to residues Cys54 and Gln55.

Figure 2.6 below shows a plot of the E_{inter}, which corresponds to the sum of restraint, vdW and electrostatic energy terms for the 200 water refined complexes from the semi-flexible docking, as a function of their backbone r.m.s. deviation from the lowest energy structure. The r.m.s. deviation of the structures converged to 0.49 ± 0.07 Å at the defined protein-DNA interface. The resulting structures had an average buried surface area of 3721.98 ± 359.80 $Å^2$. After the analysis, one cluster was obtained for all calculated structures (based on a cluster minimum size of 4 structures and an r.m.s. deviation of 1.5 Å). The resulting structures in this cluster were also run through the structure quality program PROCHECK. This confirmed that 96.4% of residues in the structures were in allowed Ramachandran space or better. Structural statistics are given in Table 2.6. The lowest energy HADDOCK score structure is shown below in Figure 2.7.

Table 2.6. HADDOCK refinement statistics. [a]Pairwise r.m.s deviation was calculated (10 structures) based on all heavy atoms at the interface of binding.
[b]Those residues that are in the disallowed regions are either in loop 3 (residue Asp40) and/or at the C-terminus of the protein (residues 51-55).

Violations (mean and S.D.)	
Distance constraints (Å)	0.23 ± 0.04
Deviations from idealized geometry	
Bond Lengths (Å)	0.0031 ± 0.0001
Bond Angles (°)	0.64 ± 0.01
Impropers (°)	0.51 ± 0.01
Average pairwise r.m.s[a] (Å)	
Interface	2.84 ± 0.64
Buried surface area ($Å^2$)	3721.98 ± 359.8
Ramachandran analysis	
Favored	71.5 %
Allowed	24.8 %
Generously allowed	1.1 %
Disallowed[b]	2.6 %

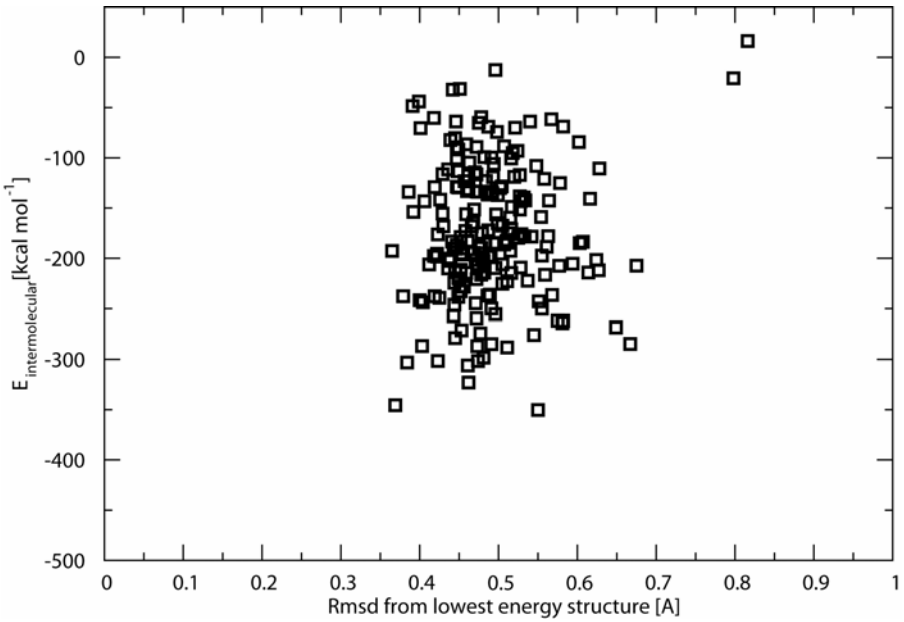


Figure 2.6. HADDOCK score versus r.m.s.d for the AbrBN[55]-tetramer in complex with *abrB8* (semi-flexible docking protocol). The results show one cluster of structures with an r.m.s.d range of 0.35 to 0.7 Å.

The calculated AbrBN[55]-tetramer/*abrB8* structures are very similar with respect to protein fold. However, the calculated *abrB8* conformations (in Table 2.7 below) are more variable. Analysis of the base pair and base pair step parameters of *abrB8* showed a deviation in the following B-DNA parameters: twist, roll, tilt, opening, propeller and buckle. Previous studies have observed variations for these parameters in both free and unbound DNA when the DNA is bending and twisting [112-116]. Of all DNA parameters that spatially relate complementary base pairs and sequential base-pair steps, buckle, propeller, roll, twist and tilt are less restricted by the conformation of the sugar-phosphate backbone and Watson-Crick hydrogen bonds [109]. The variations in these parameters resulted in an overall loss of B-DNA conformation in the docking solutions as assessed by 3DNA. These solutions had

similar HADDOCK scores to structures with a smaller variation in these parameters.

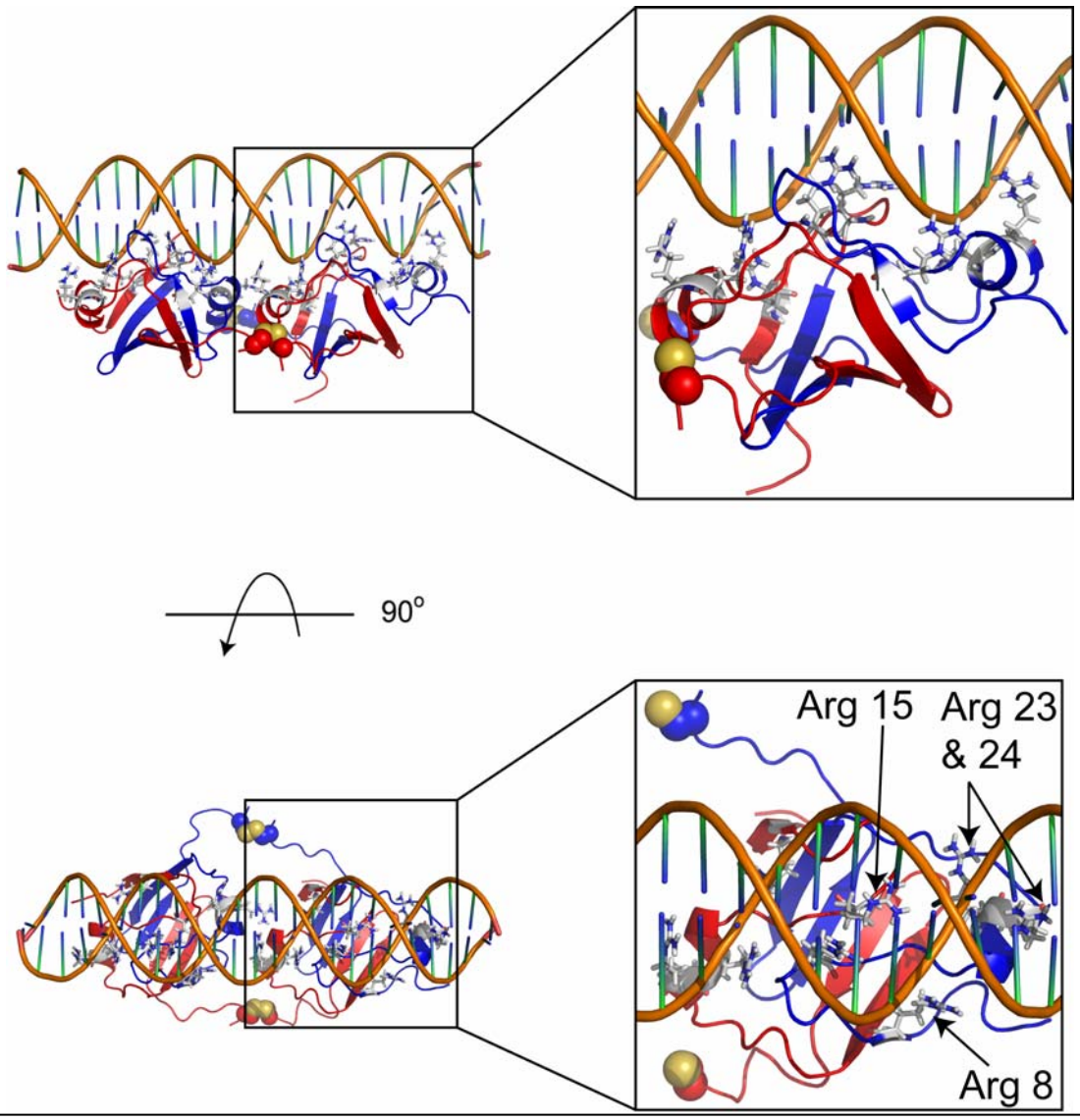


Figure 2.7. Lowest energy structure from semi-flexible docking. Two views of the lowest energy HADDOCK structure from the semi-flexible docking. Disulfide linkages are shown as spheres in the image. Insets show a detailed look at the positioning of the arginine and lysine residues involved in binding (residues labeled for monomer drawn in red).

Table 2.7. Average DNA base pair and base pair step parameters prior and post docking. *a* The average values for the canonical B-DNA input structure are shown in the left column between brackets next to each parameter. *b* Standard deviations are shown as subscripts. *c* Values from the lowest HADDOCK score structure. *d* Average values.

Parameters[a,b]	Prior – docking	Post – docking[c]
Twist ($35.9°_{0.9}$)	$36.0°_{0.7}$	$36.63°_{1.96}$
Roll ($-0.2°_{2.3}$)	$1.7°_{0.2}$	$-0.06°_{3.85}$
Tilt ($0.0°_{0.1}$)	$-0.01°_{0.48}$	$-0.13°_{2.74}$
Rise (3.4 $Å_{0.0}$)	3.4 $Å_{0.0}$	3.40 $Å_{0.14}$
Slide (0.3 $Å_{0.2}$)	0.46 $Å_{0.03}$	0.56 $Å_{0.51}$
Shift (0.0 $Å_{0.1}$)	0.00 $Å_{0.02}$	-0.08 $Å_{0.44}$
Opening (-3.3 $Å_{2.5}$)	-1.68 $Å_{0.28}$	1.86 $Å_{3.45}$
Propeller ($-10.2°_{7.3}$)	$-10.2°_{7.3}$	$3.15°_{2.29}$
Buckle ($0.1°_{0.1}$)	$0.05°_{0.56}$	$-0.23°_{2.62}$
Stagger (0.1 $Å_{0.0}$)	0.11 $Å_{0.04}$	-0.33 $Å_{0.10}$
Stretch (-0.1 $Å_{0.0}$)	-0.12 $Å_{0.04}$	0.05 $Å_{0.08}$
Shear (0.0 $Å_{0.1}$)	0.0 $Å_{0.1}$	0.04 $Å_{0.26}$
Minor groove width[d]	5.9 Å	6.5 Å
Major groove width[d]	11.4 Å	10.5 Å

Discussion

Investigation of Structural and Dynamic Characteristics with Implications for DNA Binding Specificity

In order to further our understanding of DNA recognition and binding by the 'AbrB-like' transition-state regulators, and in particular the three-fold nature of DNA-binding exhibited by AbrB, we have extended previous structural studies of AbrBN and AbhN. In this present work we conduct structural and dynamic comparisons between the RDC refined structures of AbrBN, AbhN, and the DNA-binding domain of a third AbrB-like transition-state regulator, SpoVTN, and devise a structural model for the mode of interaction between an AbrBN and a DNA target [28, 35].

The similarities in secondary structure and global fold arrangement between the structures of AbrBN and AbhN noted in our previous studies are conserved in the RDC refined structures calculated in this work (Figure 2.1A and B, respectively). The similarities between these two proteins and the RDC refined structure of SpoVTN are equally abundant. Previous work has shown that the N-terminal DNA-binding domain of SpoVT is dimeric [43, 44] and as shown in Figure 2.3A, the SpoVTN dimer is composed of monomers consisting of four antiparallel β-strands (residues 5 – 9, 15 – 17, 34 – 39, and 43 – 47) and a single α-helix (residues 20 – 26), identical to the structures of AbrBN and AbhN. SpoVTN also exhibits the domain swap fold seen in the previous structures, in which β-strands 1 and 2 from monomer A interact with β-strands 3 and 4 from monomer B (residues in this monomer will be identified by ′

notation from this point onward, e.g. Arg-15 and 15′). SpoVTN also possesses an extensive β-sheet scaffold composed of β-strands 3, 4, 3′, and 4′, as well as an extensive dimerization interface between β-strands 4 and 4′ and between β-strands 2 and 2′.

Particularly important in this dimerization interface are hydrogen bonds between residues 44 and 46 in β-strand 4 and residues 46′ and 44′ in β-strand 4′, as well hydrogen bonding between residue 16 in β-strand 2 and residue 16′ in β-strand 2′. NOE interactions between residues 44 and 46 in β-strand 4 to residues 16′ and 18′ in β-strand 2′ further define the extent of the dimerization interface. A glycine at position 42 in SpoVT (not present in AbrB or Abh) extends the length of loop 3 and shifts the register of β-strand 4 to residues 43 – 47 as compared to residues 41 – 46 as seen AbrBN and AbhN. This change in register also changes the center of symmetry from residue 44 (and 44′) as seen in AbrBN and AbhN to residue 45 (and 45′) in SpoVTN. As seen below in Figure 2.8, the RDC refined solution structures of AbrBN, AbhN, and SpoVTN are extremely similar, as evidenced by r.m.s. deviations of 1.408 Å between AbrBN and AbhN, 2.096 Å between AbrBN and SpoVTN, and 1.988 Å between AbhN and SpoVTN (backbone Cα atoms for residues 4 – 47 for all three proteins). Due to the fact that the β-sheet is located distal to the putative DNA-binding surface, the slight difference noted in the register of β-strand 4/4′ most likely does not have implications for directing DNA target specificity between the three proteins.

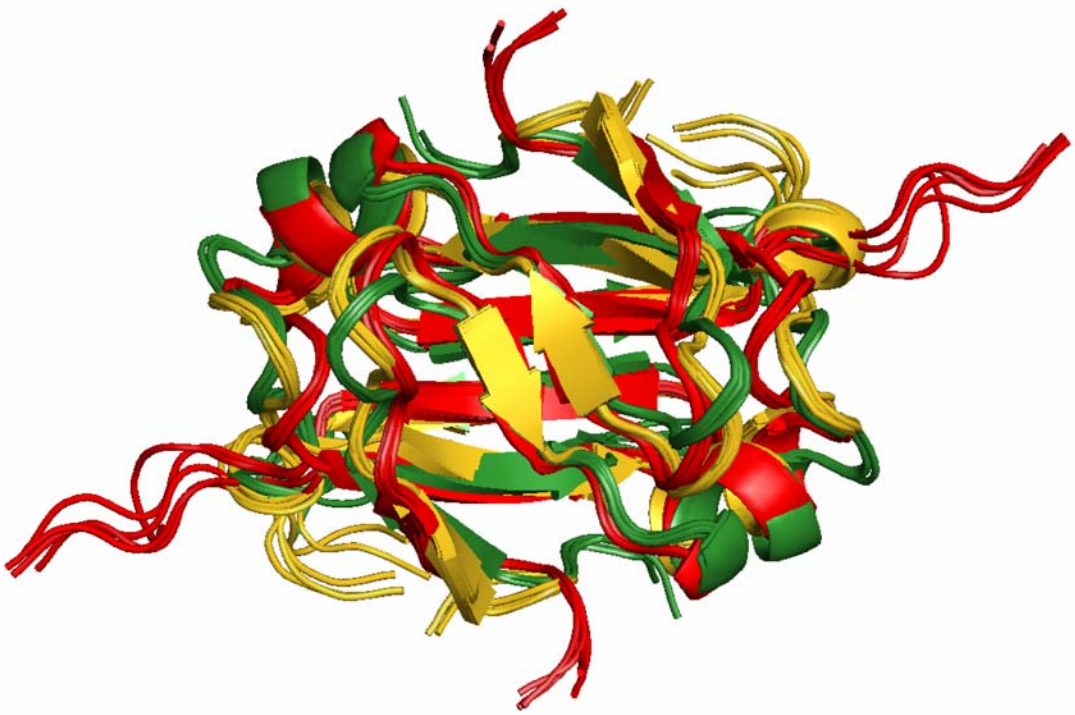


Figure 2.8. Structural overlay of the RDC-refined structures of AbrBN, AbhN, and SpoVTN. AbrBN is shown in red, AbhN is shown in gold, and SpoVTN is shown in green. The orientation of these molecules is identical to that shown in Figure 2.1A, 2.1C, and 2.3B.

The α-helices, containing residues critical to the DNA-binding properties of these proteins display significant similarities. In each of the proteins, the α-helix is composed of residues 20 – 26. While the previous study noted that the angle of the α-helix with respect to the β-strand 2 differed by approximately 2.9^{o}, the RDC refined structures revealed a greater difference in this angle. The difference in this angle between AbrBN and AbhN is 9.0^{o}, 21.2^{o} between AbrBN and SpoVTN, and between AbhN and SpoVTN this angle is 12.2^{o}. Yet as is

the case in the previous study, the pitch of the α-helices with respect to one another is very similar. This similarity in pitch most likely has functional relevance to the general recognition properties of these DNA binding proteins, while the dissimilarity in angle between the structures of AbrBN, AbhN and SpoVTN most likely has implications for the target specificity of each.

Mutagenic analysis of AbrB has identified Arg-8, Arg-15, Arg-23, and Arg-24 as residues critical to AbrB's ability to bind DNA [10, 16, 18]. In our previous study, these residues were found to occupy identical positions in AbrBN and AbhN. Extending this analysis to the RDC refined structures of AbrBN, AbhN, and SpoVTN very slight differences are noted (Figure 2.9).

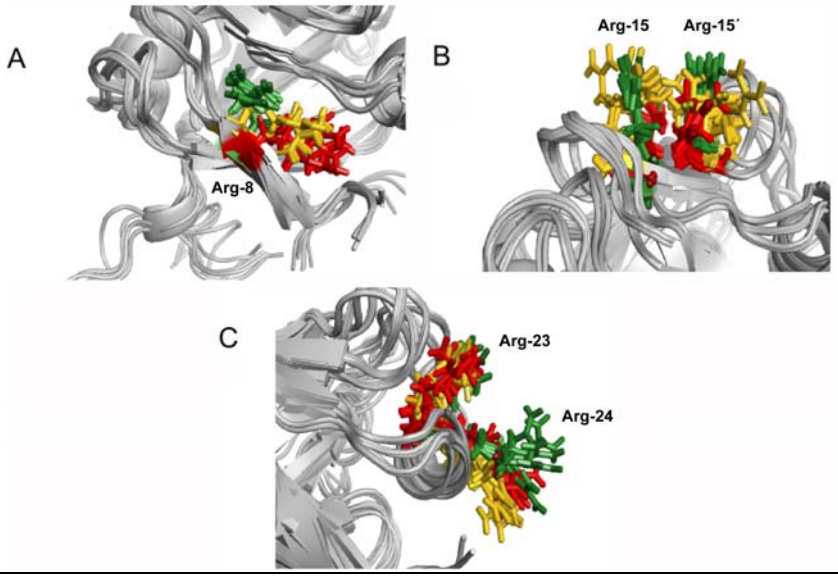


Figure 2.9. Spatial arrangement of arginine residues critical in DNA-binding by AbrBN, AbhN, and SpoVTN. Backbone overlay of ensemble solution structures of AbrBN, AbhN, and SpoVTN showing the orientation of sidechains of important arginine residues. *A* Arginine 8, *B* Arginine 15, *C* Arginines 23 and 24. Residues for AbrBN are colored red, residues for AbhN are colored gold, and residues for SpoVTN are colored green.

Looking at Arg-8 in Figure 2.9A, this residue has nearly identical positioning and orientation between AbrBN and AbhN, while in SpoVT this residue is oriented towards the loop 1. Arg-15 positioning and orientation is very similar between AbhN and SpoVTN while in AbrBN this residue appears to cross over the dimerization interface between β-strand 2 and 2′ (Figure 2.9B). Finally, Arg-23 and Arg-24 appear to be nearly identical between these three proteins (Figure 2.9C). This similarity in the spatial arrangement of Arg-23 and Arg-24 most likely has functional relevance to the general recognition (promiscuous) properties of these DNA binding proteins, while the dissimilarity in the spatial arrangement of Arg-8 and Arg-15 between the structures of AbrBN, AbhN and SpoVTN most likely has implications for the target specificity of each.

Figure 2.10 shows the electrostatic surface potential of the putative DNA-binding surface of AbrBN, AbhN, and SpoVTN, respectively. The orientation of these molecules is the same as that shown in Figure 2.8. Each of these plots has a considerable amount of positive electrostatic potential, arising primarily from the conserved arginine and lysine residues (Arg-8, Arg-9/Lys-9, Arg-15, Arg-23, Arg-24 and Lys-31), which as mentioned previously, have been identified as critical for the ability of AbrB to bind DNA. As was noted in our previous study of the non-RDC refined structures of AbrBN and AbhN, Arg-23 and Arg-24 continue to display similar electrostatic surfaces in the three RDC refined structures presented here. In Figure 2.10B and C we note that the electrostatic character provided by Arg-8, Arg-15, and loop 1 is very similar between AbhN and SpoVTN, while there is a slight difference in the electrostatic surface provided by Arg-15 in AbrBN (Figure 2.10A) when

compared to the previous work. The protonation state of Glu-30 in loop 2 of AbrB has been implicated to play a role in the differences of the pH dependant binding profile of AbrB and Abh, as it is the residue in the DNA-binding surface that is different between the two proteins. These similarities and differences in electrostatic characteristics can contribute to the general and specific, respectively, mechanisms of DNA binding.

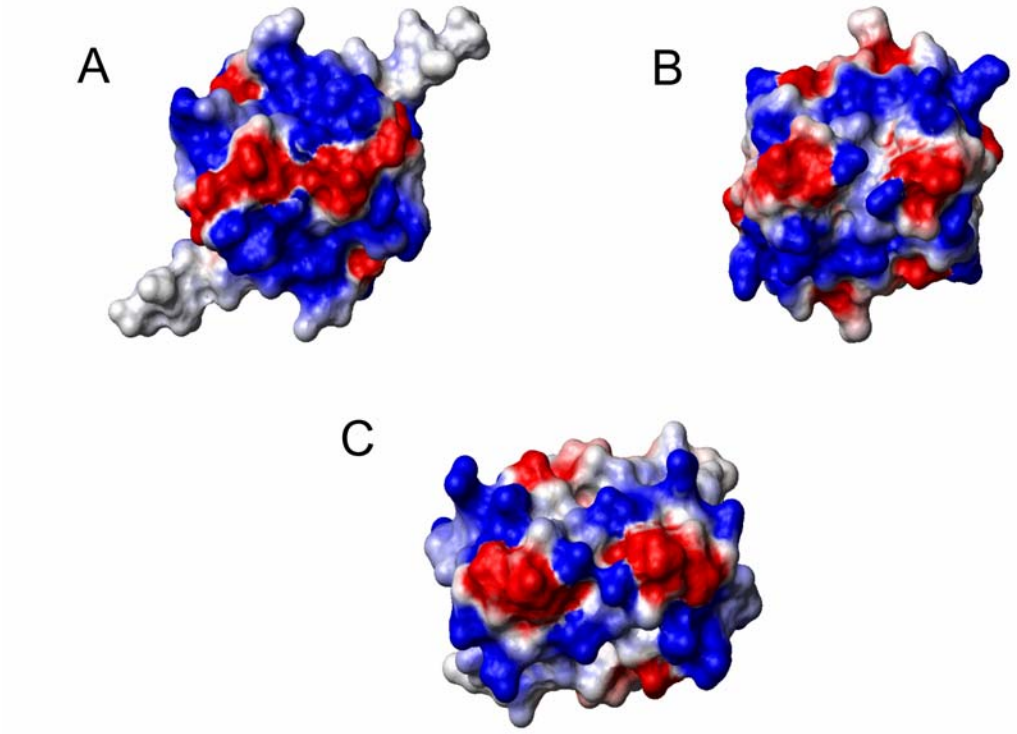


Figure 2.10. Electrostatic surface potential plots of the RDC-refined transition-state regulators. *A* AbrBN, *B* AbhN, and *C* SpoVTN. Molecular orientations are the same as shown in Figure 8.

Previous studies have suggested that loops 1 and 2 play an integral part in DNA recognition for this class of proteins as a hinge mechanism for β-strand 2 and the α-helical regions [19, 35]. Measuring the backbone Cα distance between residue 12 in loop 1 and residue 30 in

loop 2′, it is readily apparent that in AbrBN loop 1 and loop 2′ are positioned closely together (3.55 Å), while they are positioned further apart in AbhN (4.86 Å) and SpoVTN (5.45 Å) (Figure 2.11). Though the relative spatial orientation of loop 1 is very similar between all three proteins, there is a light positional shift of 1.32 Å between AbrBN and AbhN, 2.39 Å between AbrBN and SpoVTN, and 2.23 Å between AbhN and SpoVTN. Looking closely at the region of loop 2, it can be seen that loop 2 in SpoVTN is spatially positioned more closely to AbhN than to AbrBN. As measured by the distance between the Cα atom of residue 30 in each protein, the distance between loop 2 in AbrBN and AbhN is 3.69 Å, 3.06 Å between AbrBN and SpoVTN, and between AbhN and SpoVTN this distance is 1.17 Å. Despite the more similar spatial positioning of loop 2 in AbhN and SpoVTN, the orientation of loop 2 in SpoVTN is clearly more similar to that of loop 2 in AbrBN. Standing in contrast to the similarity of orientation between loop 2 in AbrBN and SpoVTN, loop 2 in AbhN is rotated by approximately 67^{o} with respect to the orientation of loop 2 in AbrBN.

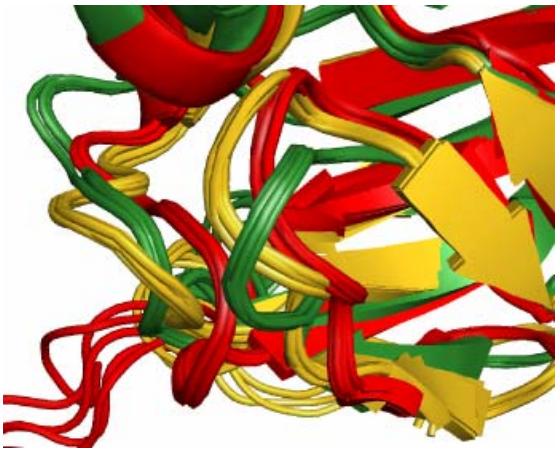


Figure 2.11. Spatial arrangement of loops 1 and 2 in AbrBN, AbhN, and SpoVTN. Backbone overlay of ensemble solution structures of AbrBN, AbhN, and SpoVTN showing the orientation of loops 1 and 2. AbrBN is colored red, AbhN is colored gold, and SpoVTN is colored green.

A further point to explore in the discussion of the loop regions in AbrBN, AbhN, and SpoVTN is the GD box in loop 2. The GD box is an evolutionarily conserved sequence motif which has been described as a hallmark of the βαβ element of the cradle-loop metafold of which the AbrB-like transition state regulators of *B. subtilis* are members [32, 117]. This sequence motif has a key structural role, forming β hairpin turns that are anchored to hydrophobic core of the protein by flanking hydrophobic residues. The glycine residue in this motif occupies the i+2 position of the β hairpin, and the NH group of the following residue (an Asp at position i+3) is hydrogen bonded to the CO group of the i residue. Further hydrogen bonds are made between the backbone CO group of the Asp residue and a backbone NH group of a residue at the end of the β1 strand. Although these characteristic hydrogen bonds were not experimentally observed in our previously published structure of AbrBN, they were observed in an independently solved structure of AbrBN [28, 32]. The RDC refined structure of AbrBN supports the possibility of hydrogen bonds being made in the evolutionarily conserved manner, as NOEs describe a structure that brings the necessary hydrogen donor and acceptor groups into sufficient proximity. Analysis of the GD box region of AbhN, on the other hand, indicates that this protein does not contain the characteristic hydrogen bond network. While the sequences of loop 2 are nearly identical between AbrBN and AbhN (aside from residue 30 which is Gln in AbrBN and Ile in AbhN), backbone hydrogen bond donors and acceptor groups are not in close enough proximity to support the characteristic hydrogen bonds. Experimentally, this observation is supported by a lack of hydrogen exchange protection for residues in loop 2. However, much like AbrBN, the loop 2 region of SpoVTN supports the classical GD box motif. Not only is the GD

sequence strictly conserved in SpoVTN, whereas the Gly residue is replaced by a Lys residue in both AbrBN and AbhN, but both hydrogen exchange and NOE structural information supports the classical GD box hydrogen bond network, forming a β hairpin turn and linking loop 2 to strand β1 via a hydrogen bond between Asp32 and Ile10. The GD boxes of the loop 2 regions of AbrBN (both the RDC refined and independently solved structures), and the RDC-refined structures of AbhN and SpoVTN are illustrated below in Figure 2.12. Experimental hydrogen exchange data for the amide proton groups of residues 10 and 32, those participating in conserved GD box hydrogen bonds, in AbrBN, AbhN and SpoVTN are shown below in Figure 2.13.

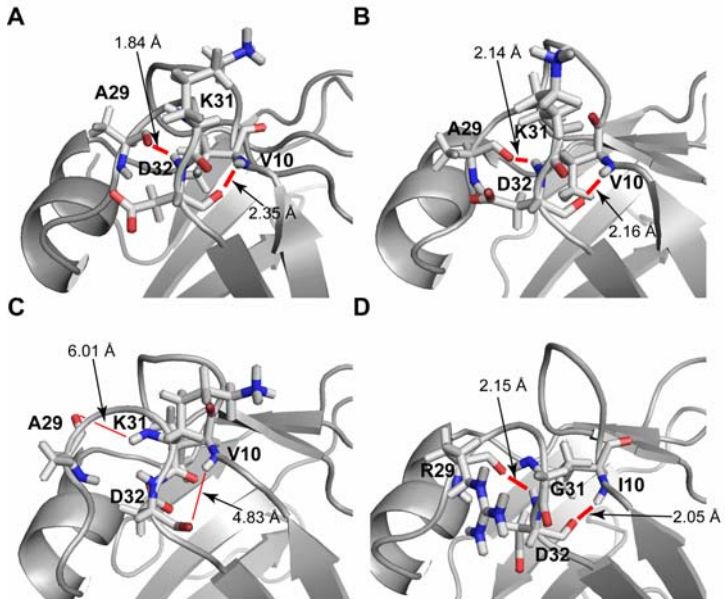


Figure 2.12. GD box motif of transition-state regulators. The residues involved in the characteristic hydrogen bond network of the GD box are shown in the above structures. *A* the GD box motif of loop 2 in the RDC-refined structure of AbrBN, *B* the GD box of the AbrBN structure by Coles *et al.*, *C* the GD box of the RDC-refined structure of AbhN, and *D* the GD box of the RDC-refined structure of SpoVTN. Key residues involved in the GD box hydrogen bonding network are depicted and corresponding CO – HN distances are noted. Thick red lines indicate probable hydrogen bonds, thin red lines indicate improbably hydrogen bonds.

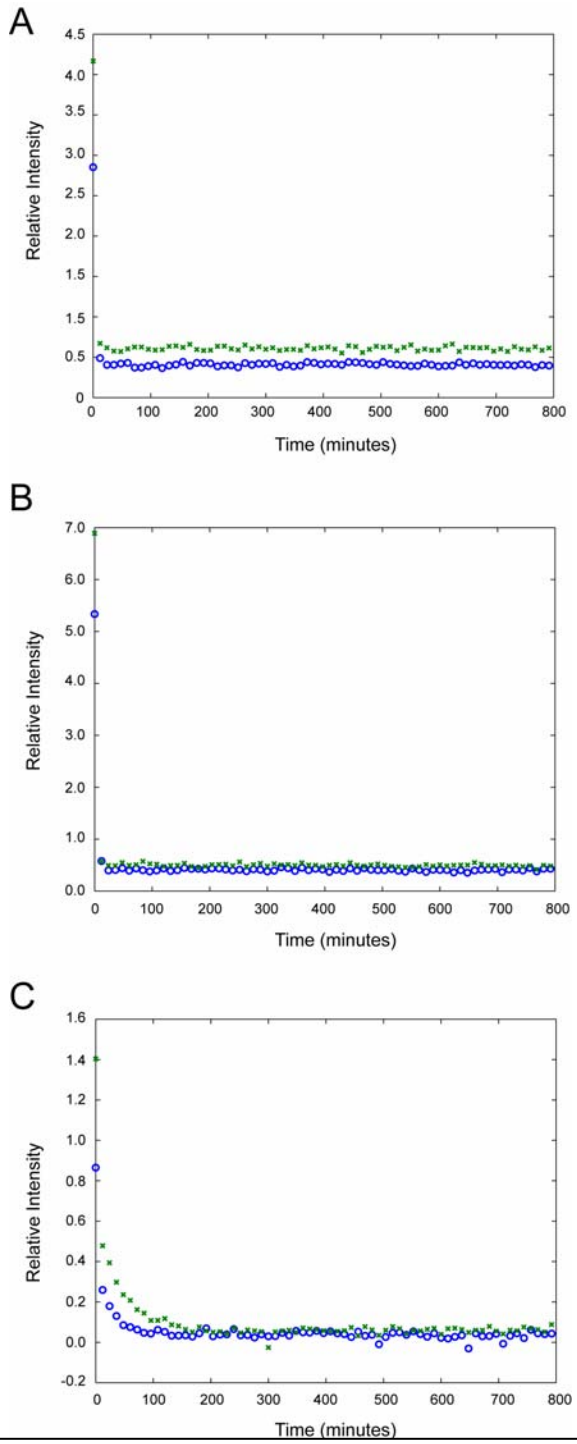


Figure 2.13. Hydrogen exchange data for the "GD-Box" region of AbrBN, AbhN and SpoVTN. 66 sequential 12 minute ^{1}H-^{15}N HSQC were recorded to determine exchange protected amide protons. A, AbrBN, B AbhN and C, SpoVT. "o" is residue 10 and "x" is residue 32.

It is worth further exploring the differences noted above in the GD-box region of the transition-state regulators and their relevance to the functional activities of DNA-recognition and binding. It has been emphasized that the GD-box is a defining structural characteristic found in all members of the cradle-loop metafold, with a key structural role as outlined above. However, given the three-fold nature of DNA-recognition and binding activity described previously in the transition-state regulator proteins, it is critical that these proteins exhibit differing properties in adhering to the presumed canonical GD-box. And given that loops 1 and 2 have roles in DNA interactions, not only via direct interactions with the DNA molecule but also in restricting the conformations of critical arginine residues, it is crucial that these proteins have different characteristics in this region, allowing them to recognize, with varying specificity, their set of cognate promoters. In summary, the GD-box motif appears to be evolutionarily conserved in this class of proteins, but it's robustness is variable – sometimes it is very pronounced, sometimes it is extremely loose, almost non-existent. This variability is likely critical to DNA target recognition.

To assess more accurately the correlation between looped-hinge/DNA recognition helix mobility and protein function between the three proteins, ^{15}N NMR relaxation experiments were performed. The results provide one of the most compelling examples of similarities and differences in functional mobility identified for this class of DNA binding proteins. We define dynamic motion as a measure of the propensity for residues to adopt multiple conformations. A detailed look at the S^2 general order parameter from Figure 2.4A reveals

that all three proteins follow the same trends in the mobility of backbone amide groups, including increased mobility in loop 1, β2, and the region of loop 2. These motions are critical to allowing secondary structure elements, including β2 and the α-helix, discussed further below, to reorient during complex formation with promoter DNA sequences thereby allowing the arginine residues critical to DNA-recognition and binding to make their proper contacts. This dynamic character, common to these three proteins, must be part of the general DNA-recognition properties for the transition-state regulators. However, a closer look at the S^2 values for several regions of AbrBN, AbhN, and SpoVTN reveals propensities for dynamic motion that may play a role in the specificity exhibited by each protein. For instance, AbhN is less restricted in β1, surrounding Arg8 than either AbrBN or SpoVTN, indicating a greater ability for AbhN to reorient this portion of the protein. On the other hand SpoVTN and AbrBN are less restricted in loop 1 than AbhN, especially around residue 12. In the α-helix, AbrBN and AbhN are nearly identical, but SpoVTN is slightly more restricted in motion in the N-terminal portion of the helix (including residues 21 and 22) indicating that the helix in SpoVTN is slightly less able to adopt multiple conformations, which would have consequences for complex formation. In the region of the GD-box and loop 2, both AbrBN and SpoVTN are slightly less restricted than AbhN. This, coupled with the greater mobility in loop 1 for both proteins, indicates that AbrBN and SpoVTN experience concerted motions between loops 1 and 2, which may be due, in part, to the hydrogen bonding network of the GD-box, which is completely lacking in AbhN. Finally, in the region of loop 3, SpoVTN is slightly less restricted than either AbrBN or AbhN, most likely owing to the presence of the extra glycine in this loop that is not present in AbrBN or AbhN.

Dynamic motion on similar timescales can define networks of residues within regions of a protein that are important for protein function. The clustered patterns of backbone dynamics identified in all three proteins, suggests that the ability of this class of proteins to recognize DNA sequences with no apparent consensus sequence may be attributed to a synchronization of motion between structurally independent regions of the protein. As is apparent in Figure 2.4B, the residues with chemical exchange relaxation terms (R_{ex}) and thereby exhibiting motions on the micro- to millisecond (μs-ms) timescales are rather different, yet commonalities are also readily seen. All three proteins exhibit coordinated motions on this timescale between residues in loop 1, β2 and loop 2, critical in allowing reorientation of this region, as well as the α-helix, upon complex formation. In this manner, the above motions must be part of the general recognition properties of the transition-state regulators. However, once again differences are apparent in each of these proteins, providing further insights in the nature of the specificities of each. The residues exhibiting μs-ms timescale motions are more numerous in both AbrBN and AbhN than SpoVTN indicating the presence of a more complicated network of concerted motions. AbhN has a large number of residues with these motions clustered in the N-terminal portion of the protein, especially in β1 and the loop 1 region, whereas AbrBN has several residues clustered in loop 3 that have significant motions on this timescale. The differences noted here will obviously have functional implications for the specific DNA-binding activities for each of these proteins. Those residues exhibiting concerted motions on the μs-ms timescales are plotted on the RDC-refined TSR structures below in Figure 2.14.

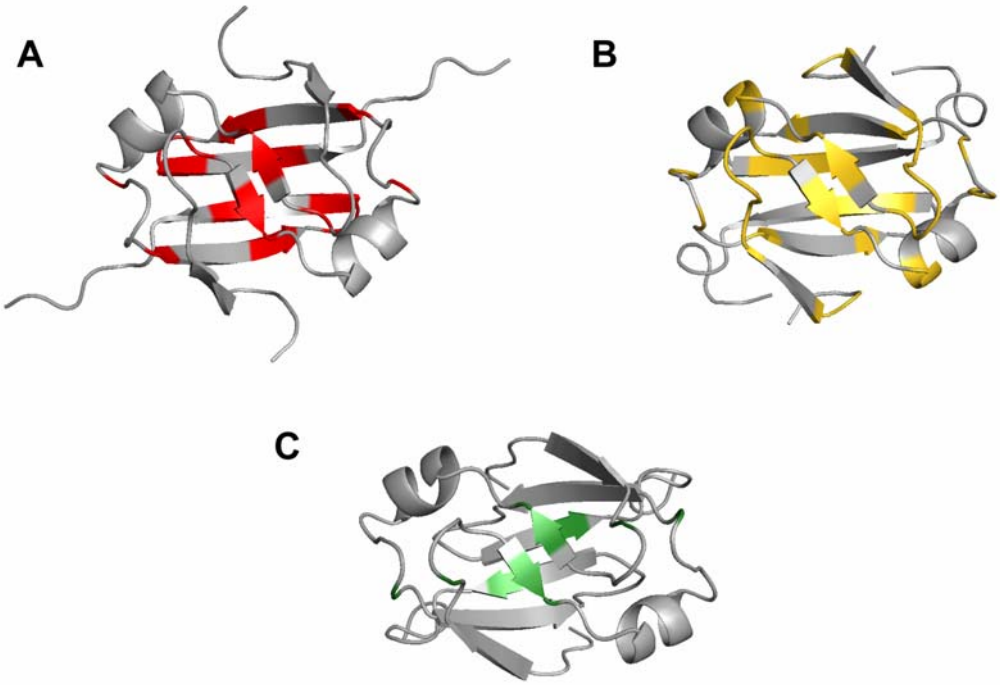


Figure 2.14. Residues experiencing motions on the micro- to millisecond timescale. *A* Residues of AbrBN with R_{ex} chemical exchange terms are shown in red. *B* Residues of AbhN with R_{ex} chemical exchange terms are shown in gold. *C* Residues of SpoVTN with R_{ex} chemical exchange terms are shown in green.

Elucidation of AbrB:abrb8 Complex

The data presented in this manuscript reveal similarities and differences that may contribute to the nature of general and specific binding by these proteins, yet the structure of an 'AbrB-like' protein bound to DNA remains elusive. The modeling of the AbrBN^{55}-tetramer-*abrB8* complex described above represents the first model of any 'AbrB-like' transition-state regulator bound to a DNA target. This model was created using all known biochemical data mutational data. Comparison of the modeled structure of the AbrBN^{55}-tetramer-*abrB8* complex generated by HADDOCK and the experimentally derived structure of unbound

AbrBN^{53} indicates that AbrBN^{55}-tetramer maintains a similar global fold upon binding (Figure 2.15). The overall r.m.s. deviation between one AbrBN^{55} dimer of the calculated structures and the solution structure of the unbound AbrBN^{53} is 2.82 Å (residues 1-53). The N-terminal dimerization interface is preserved, with minimal changes observed in the positions of β-strand 3 and β-strand 4. This observation is consistent with the role of this region of the protein being responsible for dimerization. However, upon complex formation, the α-helix and loops 1 and 2 of the AbrBN^{55}-tetramer are somewhat displaced (Figure 2.15). These regions cooperatively migrate towards the DNA phosphate backbone, while making limited base pair specific interactions. This co-operative migration complements known dynamics data that suggest a notable ability to change conformation in these regions – see above. The model also suggests local deformation of the DNA upon binding.

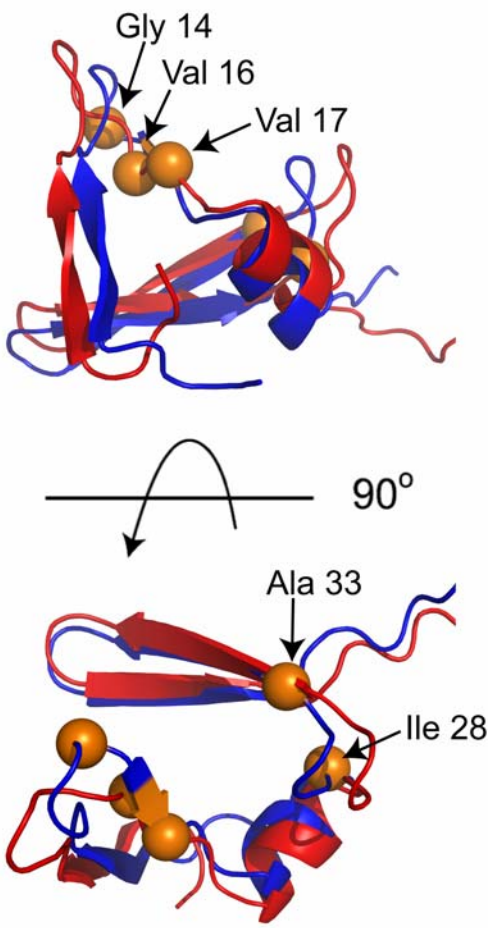


Figure 2.15. Overlay of unbound and bound AbrBN. The red ribbon structure is the lowest HADDOCK score structure from the modeling studies and the blue ribbon structure is the solution structure of the unbound AbrBN[53]. Cα atoms of residues exhibiting micro-millisecond timescale motions are shown as orange spheres. Only one monomer (plus labels) is shown for clarity. The monomeric structures shown overlay with a r.m.s.d of 1.62 Å.

In the AbrBN[55]-tetramer-*abrB8* structures calculated, β-strand 1 moves slightly, allowing Arg-8 and Lys-9 to bind to the DNA backbone while still providing an appropriate dimerization interface. Loops 1 and 2 undergo significant movement upon complex formation. In the structural model these loops move to facilitate limited contacts between Lys-31 (loop 2) and the DNA backbone. The α-helix moves towards the DNA backbone,

making contacts between Arg-23 and Arg-24 and the phosphate backbone; several hydrogen bonds are observed between the δ–guanidino groups of Arg-23 and Arg-24 and the oxygen atoms of the phosphate backbone. Arg-24 appears to undergo the most significant conformational change of the residues directly responsible for binding DNA.

A detailed view of the interface between the AbrBN^{55}-tetramer and *abrB8* (Figure 2.7 above) reveals that an extended network of intermolecular hydrogen bonds stabilizes the complex. These hydrogen bonds involve both charged and polar amino acids at the interface and within the protein structure. The structures show the positioning of the AbrBN^{55}-tetramer in consecutive major grooves (left hand panels of Figure 2.7) with the loop regions extending into the major groove to make base-pair specific interactions, primarily targeting guanine bases. For example, the δ–guanidino groups of Arg-15 (β-strand 2) form hydrogen bonds and/or non-covalent interactions with guanines (right hand panels of Figure 2.7). Additionally, the model allows for the δ–guanidino groups of Arg-8 (β-strand 1), Arg-23 and Arg-24 (both α-helix) to make contacts with the phosphate backbone (approximately 4 base pairs in length).

Analysis of the most populated hydrogen-bond interactions made between the protein and the DNA reveal extensive hydrogen bonds between the δ–guanidino groups of β-strand 2's Arg-15 and N7 and O6 of guanine residues (in the loose conserved sequence) as well as with atoms on the DNA phosphate backbone (O#P, O5' and O3'). In all models the δ–guanidino

groups of Arg-8 (β-strand 1), Arg-23 (α-helix) and Arg-24 (α-helix) and the amino groups of Lys-9 (β-strand 1) and Lys-31 (loop 2) move toward the DNA phosphate backbone and were only ever found to contact atoms on the DNA phosphate backbone (O#P, O5' and O3'). In some instances, the terminal oxygen (OE1) of Glu-12 (loop 1) was observed to have hydrogen bonds with the NH_2 groups of adenine and cytosine (H6# and H4#, respectively) while the terminal oxygens (OD1 and OD2) of Asp-11 (loop 1) were observed to have hydrogen bonds with the NH_2 groups of cytosine (H4#).

Millisecond timescale motions (see above) were observed for residues in loop 1, loop 2, and β-strand 2. Residues Gly-14 (loop 1), Ile-28 and Ala-33 (loop 2) have millisecond timescale motions and bracket loops 1 and 2. These residues provide a convenient hinge mechanism that allows loops 1 and 2 to reorient and make hydrogen bonds with the phosphate backbone. Residues Val-16 (β-strand 2) and Val-17 (β-strand 2) also have motions on the millisecond timescale providing a mechanism in which Arg-15 (β-strand 2) is allowed to make specific base pair interactions within the major groove. The motions of loops 1 and 2 also allow the α-helix to alter its conformation and migrate towards the DNA backbone (Figure 2.12). The propensity of these regions to readily adopt multiple conformations enables the straightforward re-positioning of residues critical to the binding interaction (red ribbon structure) (Arg-8, 15, 23, 24 and Lys-9 and 31).

In conjunction with minor structural changes taking place in the $AbrBN^{55}$-tetramer, the DNA target also undergoes limited deformation from canonical B-DNA. However, it retains

torsion angles and structural parameters that are well within the allowed values of known protein-DNA structures and X-ray crystallographic studies of B-DNA [109, 118, 119]. This observation is consistent with previous studies showing that AbrB can bind both bent and non-bent DNA, and can induce DNA bending upon binding [34].

The specific protein:DNA interactions described can be accommodated within three broader binding models *in vivo*. These three models are shown schematically in Figure 2.16. Electrophoretic mobility assays demonstrate that AbrB can bind both bent and linear DNA [34]. Figure 2.16A shows tetrameric AbrB binding to 40-50 base pair looped DNA. This model requires that the approximate coverage protected from DNase I cleavage be 60 to 70 base pairs. Although some AbrB-target interactions have been shown to result in protection regions of this length, the universality of this model is not supported by other studies which show that the base pair regions typically protected from DNase I cleavage in many AbrB DNA targets are between 25 and 45 base pairs with nanomolar binding affinity (M.A. Strauch, unpublished). Figure 2.16B shows two tetrameric AbrB molecules binding to 40-50 base pair looped DNA in consecutive major grooves. This model requires that the approximate coverage protected from DNase I cleavage would be 80 to 90 base pairs. Again, while there are some examples of interactions giving this extent of protection from DNase I cleavage, the model is not universally supported by all previous DNase I footprinting studies (M.A. Strauch, unpublished). Figure 2.16C shows two tetrameric AbrB molecules binding to ~ 40 base pair linear DNA in pairs of consecutive major grooves. DNase I footprinting studies showing 25-45 base pairs of AbrB-afforded protection support this model for the

interaction of AbrB with those sites. Additionally, previous μESI-MS studies showed that the largest detected interaction was that of a single AbrB tetramer with a single 30bp DNA fragment, affording a molecular weight of 60 kDa. This molecular weight precludes the binding of multiple DNA fragments. Model C could also apply to those interactions giving longer regions of DNase I protection if more than two tetramers of AbrB were binding side-by-side. While *in vitro* observations indicate that most AbrB-DNA interactions can be accommodated by model C, suggesting that this is the most likely binding mode, none of the three models can be ruled out entirely and each may reflect a different form of AbrB interaction that occurs in a target-specified manner.

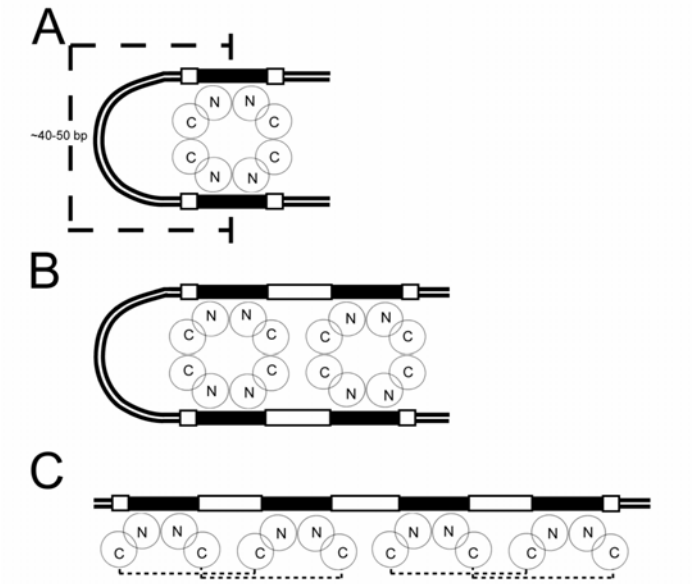


Figure 2.16. Schematic models for AbrB-DNA binding. A, binding model of one tetramer bound to two target regions separated by a DNA loop (typical length of DNA loops shown). B, binding model of two AbrB tetramers bound to four target regions separated by a DNA loop. C, binding model of two AbrB tetramers bound to four target regions *not* separated by a DNA loop. Solid black boxes denote the loose conserved recognition sequence of TGGNAN; white boxes denote 4 base pairs immediately adjacent to this sequence, "N" denotes the N-terminal domain of AbrB while the "C" represents the C-terminal domain. Small dashed lines in panel "C" define the AbrB tetramer.

Conclusions

Elucidating the mechanism by which transition-state regulators interact with DNA targets is particularly important as these proteins are critical in the ability of bacteria to adapt to changing environments through metabolic changes, toxin expression, and alternative developmental pathways. This study characterized the structure and dynamics of three such transition-state regulators, representing the first attempt to determine how a transition-state regulator binds its DNA target and what structural changes occur during this process. The calculated AbrBN^{55}-tetramer-*abrB*8 structures are especially helpful in identifying the role of residues at the protein-DNA interface. The biochemical data driven docking calculations were performed to allow flexibility in the backbone and the sidechains of appropriate residues, enabling them to alter their conformations upon complex formation. Subsequent analysis of the modeled structural ensembles also provides information on the conformational changes occurring in both protein and DNA upon binding. The implications of this study go beyond the structural determination of a protein:DNA complex. This model can now be used to determine how to structurally exert tighter regulation on DNA targets that result in spore formation, virulence gene expression, etc by this unique class of "AbrB-like" transition-state regulators.

Future Directions

Work has begun in earnest to describe the domain organization and multimeric interfaces of a full length transition-state regulator, beginning with AbrB. In order to gain insights into such information, a mass spectrometry based collaboration has been undertaken with Dr. Michael

Goshe in the Dept. of Molecular and Structural Biochemistry at NCSU. This mass spectrometric approach utilizes a novel cross-linking agent, SuDP, developed by Dr. Goshe and Erik Soderblom, to link lysine residues (both intra- and intermolecularly) within a certain distance of one another in a protein structure, thereby allowing for the determination of low-resolution structural restraints. This methodology has been demonstrated on AbrBN, and the determined distances correlated with the NMR solution structure (Figure 2.17). Furthermore, the general fold of AbrBN was established extremely rapidly using this technique, thus paving the way for this technique in analyzing the domain organization and global fold of full-length AbrB. Of particular interest of full-length is the nature of the intermolecular interfaces of AbrB, i.e. how the dimers of the homotetramer interact with one another: monomers A ↔ B and C ↔ D, or monomers A ↔ D and B ↔ C.

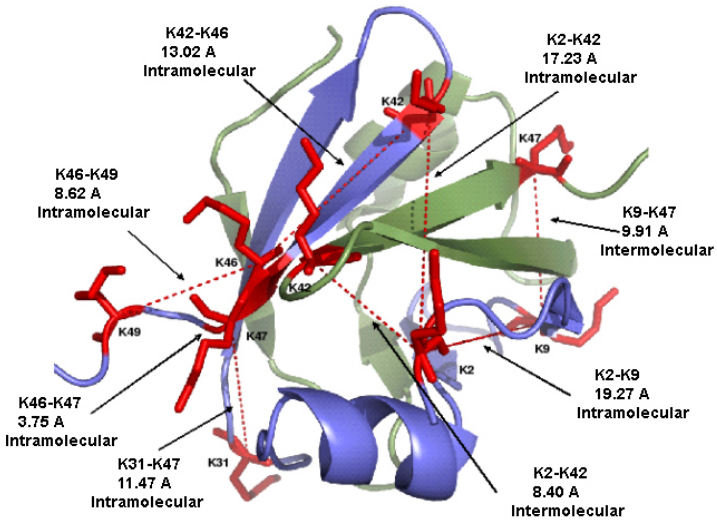


Figure 2.17. AbrBN cross-linked with SuDP. The monomeric subunits of AbrBN are colored green and blue. Cross-linked lysine residues are colored red.

Chapter 3: Sub-classification of Response Regulators Based on Hydrophobic Surface Characteristics

Abstract

The two-component signal transduction system is a ubiquitous signaling module present in most prokaryotic and some eukaryotic systems. Two conserved components, the sensor histidine protein kinase and response regulator, function as a biological switch, sensing changes in the environment, and eliciting a specific cellular response. Extensive studies have classified the histidine protein kinase and response regulator proteins using primary sequence characteristics, domain identity, domain organization, and biological function. It has been proposed that structural analysis of the surface properties of the highly conserved receiver domain of response regulators can be used to build upon previous classification methods. Previous studies of the OmpR sub-family response regulators in *Bacillus subtilis* and *Escherichia coli* revealed a notable correlation between the response regulator receiver domain surface classification and previous classification of cognate sensor histidine kinase proteins. We have extended these studies to analyze the receiver domains of all predicted response regulator proteins in the marine dwelling bacterium *Vibrio vulnificus*.

Introduction

Biologists have long used classification as a means of organizing and understanding the complex nature within and between organisms. With the advent of large-scale genomic sequencing, it becomes possible for the experimentalist to infer the function of a protein

found in a less studied organism based upon information available from the extensive studies of other model organisms, such as *Bacillus subtilis* and *Escherichia coli* – the so-called *structure-function* relationship [120]. Classification by protein *structure* broadly describes the use of the primary sequence (amino acid composition) and structural folds (three-dimensional structures of individual protein domains) to suggest possible relationships between distinct sets of proteins – in essence, the *chemical* makeup of proteins. On the other hand, classification based on protein *function* refers to the use of *biological* information, such as the role of the protein or domain within the organism.

Bacteria are able to respond rapidly to variable environments by translating an environmental signal into a cellular response by utilizing a ubiquitous signaling mechanism known as the two-component signal transduction system. Upon receiving a cue from the extracellular environment, a sensor histidine kinase autophosphorylates in an ATP-dependant manner on a conserved histidine residue. This phosphoryl group is subsequently transferred to the conserved aspartic acid binding pocket of the regulatory domain of a response regulator, thereby activating the output domain [45-53]. Despite the varied nature of cellular activity of the activated response regulator, the precision of the recognition process between the sensor histidine kinase and the regulatory domain cannot be overstated, as an imprecise interaction between a sensor histidine kinase and a non-cognate response regulator may result in an incorrect and non-adaptive response [57]. The crucial nature of this interaction is further complicated when one considers both the degree of conservation in the structures of both the histidine kinase and the response regulator, as well as the number of two-component signal

transduction systems that various bacterial species possess. Both *B. subtilis* and *E. coli* are known to possess more than 40 different two-component systems, and the draft release of the genome of *Nostoc punctiforme* has revealed the potential for greater than 100 two-component systems [37, 121, 122]. Several examples of the structurally conserved receiver domain that have been solved by both x-ray crystallography and solution NMR spectroscopy are shown below in Figure 3.1 [123-136]. Out of this inherent complexity, the central issue arises: what specifies the proper interaction between a sensor histidine kinase and its cognate response regulator?

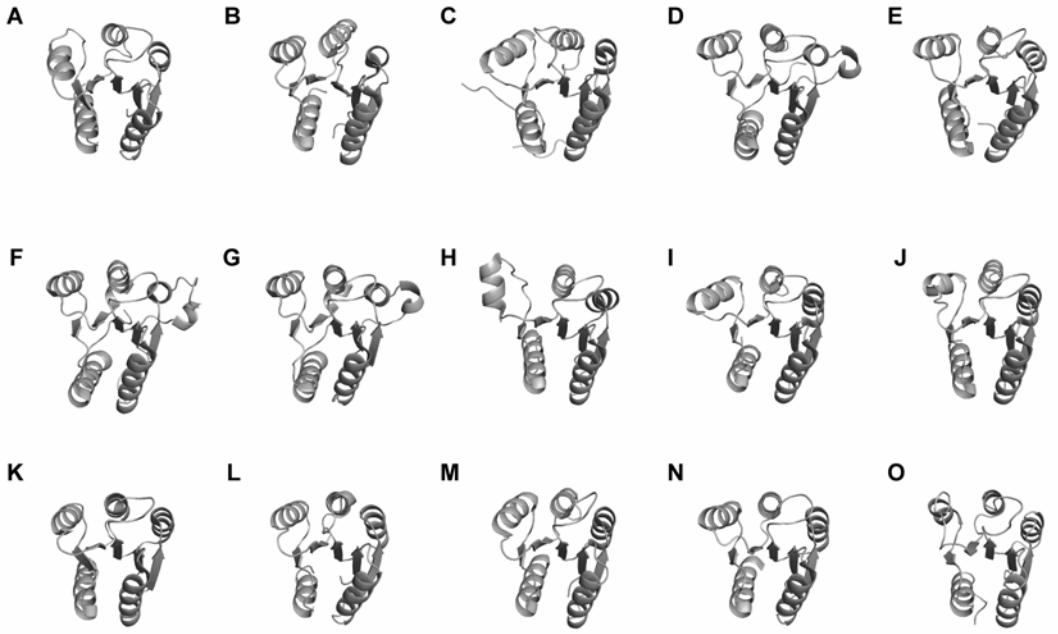


Figure 3.1. Structures of regulatory domains in response regulators. Regulatory domains of: (A) PhoB from *E. coli* (PDB:1B00), (B) FixJ from *R. melitoti* (PDB:1DBW), (C) Etr1 from *A. thaliana* (PDB:1DCF), (D) Rcp1 from *Synechocytis sp*. (PDB:1I3C), (E) CheY from *E. coli* (PDB:1JBE), (F) RcpB from *Tolypothrix sp.* PCC 7601 (PDB:1K66), (G) RcpA from *Tolypothrix sp.* PCC 7601 (PDB:1K68), (H) DrrD from *T. maritima* (PDB:1KGS), (I) DivK from *C. crescentus* (PDB:1M5T), (J) PhoP from *B. subtilis* (PDB:1MVO), (K) NtrC from *A. aeolicus* (PDB:1NY5), (L) CheY from *T. maritima* (PDB:1TMY), (M) PleD from *C. vibrioides* (PDB:1W25), (N) ArcA from *E. coli* (PDB:1XHF), and (O) Spo0F from *B. subtilis* (PDB:2FSP)

In an effort to provide some insight into the nature of specificities guiding productive interactions, two-component signal transduction systems have been a subject of classification on both the structural and functional levels in order to understand evolutionary relationships. The first formal classification of these proteins was accomplished by comparative sequence analysis of a sequence set made up of 348 histidine protein kinase domains and 298 response regulator receiver domains [49]. The analysis utilized features such as primary sequence alignments (conserved amino acid residues), tertiary structure (fold and domain organization), protein function (enzymatic, DNA-binding, ligand-binding, etc.) and biological function (involved in sporulation, chemotaxis, etc.). The study revealed that distinct sub-groups of histidine protein kinases exist and a strong correlation between the histidine protein kinase and response regulator sub-families was noted. This suggested that sub-classes of two-component systems have evolved independently. In a subsequent study, the histidine protein kinases in *B. subtilis* were analyzed and classified into groups as a basis for determining specificity of the response regulator for its cognate histidine protein kinase [137]. These studies provided a detailed classification of histidine protein kinase based on such features as sequence homology around the conserved phospho-histidine residue and the lengths of intra-cellular and extra-cellular loops and domains. However, only a broad classification of response regulators was provided from these previous studies, primarily focusing on sequence similarity of the receiver domain and cellular function of the output domain.

In an effort to create more detailed sub-classes of response regulators, information derived

from a Spo0F-Spo0B co-crystal structure has been used to analyze sequence alignments of the receiver domains of the OmpR sub-family response regulators [57, 138]. A correlation was observed between residues that comprise the interaction surface between the receiver domain of response regulator proteins and the four-helix bundle motif of histidine proteins kinases and phosphotransferase proteins. These residues consist of three general types of amino acids: i) essential invariant catalytic residues directly involved in the phosphotransfer mechanism, ii) anchor residues that establish broad orientational contacts for catalysis, and iii) recognition residues that ensure the correct two proteins come together. This patch of residues on the response regulator protein, suggested to be important for recognition specificity, is found on the surface containing α-helix 1, α-helix 5 and the β4-α4 loop. Subsequent co-crystal structures and biochemical analyses have confirmed the role of this receiver domain surface in the interaction between the response regulator and its cognate four-helix bundle motif (as illustrated below in Figure 3.2) [139-141]. This was further confirmed by another study that noted the four-helix bundle motif contains recognition specificity for its cognate response regulator [142]. All of these observations lead to the notion that the interaction surface on the response regulator is conserved, not only within a single protein amongst a variety of related organisms, but perhaps amongst all response regulator proteins.

The Hoch and Varughese study made several observations concerning the nature of the residues that make up the interaction surface [57]. Catalytic residues are almost strictly invariant and conserved, which is not surprising as the residues are the primary determinants

of the directionality and equilibrium of phosphoryl transfer, and minor changes can have extremely detrimental effects on bacterial adaptability. Anchor

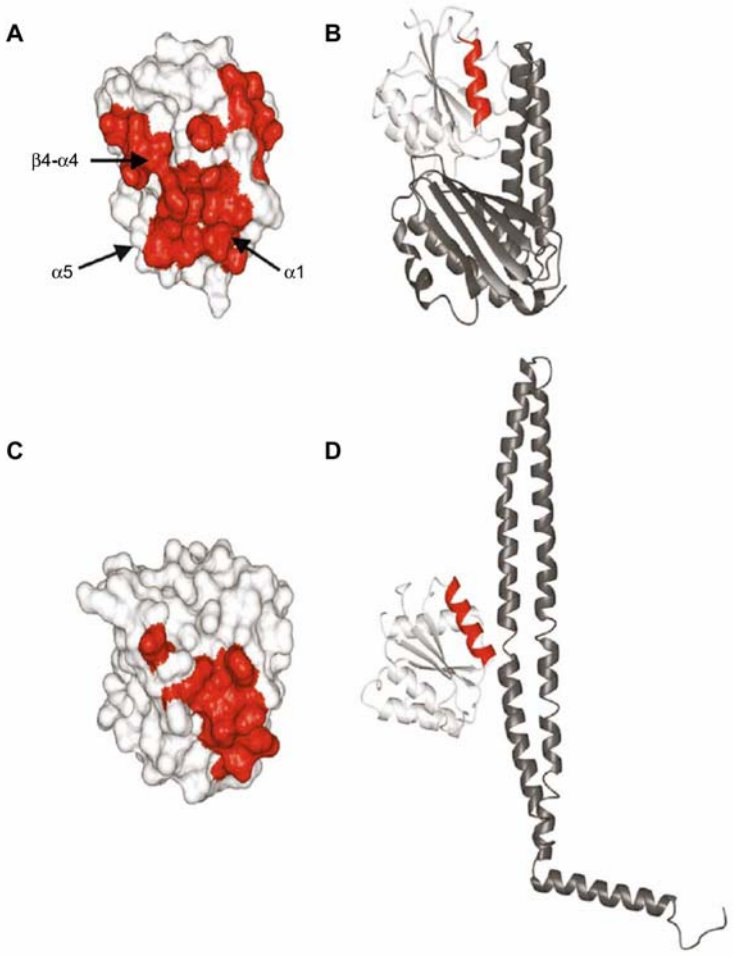


Figure 3.2. Response regulator:four-helix bundle co-crystal structures. *A* Regions of Spo0F that contact Spo0B in the co-crystal structure (shown in *B*) are colored in red. *C* Regions of CheY that contact CheZ in the co-crystal structure (shown in D) are colored in red.

residues are also conserved, although perhaps to a lesser degree, and they determine how the two protein domains will come together. However, it is in the recognition residues, that a degree of moderate variability is seen. It is the recognition residues that either allow or

prevent the two proteins to fit together by introducing charge and size constraints in the interaction surface. To ensure that unproductive interactions between the histidine protein kinase and the incorrect response regulator are prevented, a means of discrimination must be discernible. If this is the case, then it is plausible that a specific region on the receiver domain surface that shows heightened variability across the family may act as the primary contributor in histidine protein kinase interactions.

Based on the above studies, an interesting question arises: can biological information, such as the evolutionarily conserved interaction surface on response regulator receiver domains, detailed above, be used in conjunction with sequential and structural information to further sub-classify members of this related sub-family? Put another way, can the similarities and differences of the residues on the receiver domain of the response regulator that comprise the interaction surface be utilized in conjunction with structural information to further sub-classify these protein domains? The first study to investigate this premise has been conducted using the OmpR sub-family of response regulators from *B. subtilis* and *E. coli* [143]. Using a comparative modeling approach, homology models for 30 response regulator receiver domains were generated, using previously solved structures deposited in the Protein Data Bank as structural templates. The results of this study indicated that the highest degree of hydrophobic content and variation is contained in the interaction surface comprised of α-helix 1 and the α-helix 1/α-helix 5 interface. This observation was supported by previous studies that indicated that approximately 75% of the receiver domain contacts to the histidine kinase reside in this region [57, 138, 139]. The amount of hydrophobic variability present in

this interaction surface provides a means for discrimination between histidine protein kinase domains. Further validating this methodology as a means for sub-classification, there was a significant correlation between the response regulator sub-classes developed using comparative surface analysis and the histidine protein kinase sub-classes based on sequence comparison and domain architecture [49, 137].

Given the success of an analysis restricted to the OmpR sub-family of response regulators of *B. subtilis* and *E. coli* (for a total of 30 sequences), the question arose as to how successfully this methodology could be applied to larger systems, such as the full complement of response regulator receiver domains predicted to exist within a whole genome. In an effort to study the success and practicality of this approach to sub-classification in the context of a larger system, the predicted response regulator receiver domains of the pathogenic marine bacterium *Vibrio vulnificus* YJ016 were chosen as the target set [144]. With a total of 92 predicted response regulator receiver domains, the choice of the *V. vulnificus* YJ016 genome provided several advantages, first and foremost being that structural and functional annotations for this genome were available from the Superfamily server, a collection of Hidden Markov Models for each of the 1539 superfamilies in the SCOP database [145]. This greatly aided in the identification of sequences predicted to contain response regulator receiver domains.

Methods

Target Identification and Template Selection

The response regulator receiver domains used in this study were obtained using the SUPERFAMILY [146] database entries CheY-like receiver domains for *V. vulnificus* YJ016, and the respective GENBANK [147] entry was used to approximate the receiver domain sequence. The BioPython package [148] was used for semi-automation of the sequence analysis and modeling procedures. This utilized the capabilities of the built-in packages to: (i) obtain *V. vulnificus* response regulator protein sequences from the GenBank database, (ii) use the GenBank sequence records to narrow down the receiver domain sequence ranges, and (iii) feed this information for subsequent external analyses (i.e., the MODELLER program). During initial stages of implementation, it was apparent that the GenBank sequence record describing the receiver domain (Region; region_name="REC") was often inaccurate or missing, which is likely a result of incomplete genome annotation, particularly on newly sequenced genomes. Because of this, full-length sequences were obtained for each *V. vulnificus* response regulator, and the sequence range for each receiver domain was manually obtained using information provided by the SUPERFAMILY database, an initial PSI-BLAST search and visual inspection of the alignments, focusing on conserved active site residues and the length of secondary structural elements, α-helix 5 in particular. In order to identify template sequences, the PISCES program [149] was used to restrict the PSI-BLAST search to a 90% nonredundant sequence search using the database pdbaanr. Templates were ranked according to the E value returned by the PSI-BLAST search.

Target-Template Sequence Alignment and Model Building

Sequence alignments between the target sequence and template structures were automated using the SALIGN and MALIGN commands in available in MODELLER 8v2 [150]. These sequence alignments were then visually inspected to assess the quality of the alignment based on the known conserved and active site residues. Ten models target were calculated for each target sequence using default MODELLER parameters with the degree of refinement set to the very fast MD annealing 'refine 1'. These ten models were calculated using the alignment between the target sequence and the sequence of the four most homologous structures.

Evaluation of Models

The models that were generated for the *V. vulnificus* sequences were not subjected to the robust analysis conducted for the OmpR sub-family of response regulators in *B. subtilis* and *E. coli*, in which models were initially ranked based on their MODELLER defined molecular probability density function, or objective function, which encompasses the overall energy constraint violation. PROCHECK was next used to analyze Ramachandran stereochemical parameters of the models, following which ERRAT and VERIFY3D were utilized to analyze interactions between different atom types and to analyze the side-chain environment based on the solvent accessibility of the side-chain and the fraction of the side-chain covered by polar atoms [107, 151, 152]. Cα RMSD between the model and the solved structure of PhoB (PDB: 1B00) from *E. coli*, as it generally had the most sequence similarity to the OmpR sub-family receiver domains in the study, was calculated using the SUPERPOSE command in MODELLER.

The comprehensive analysis described above was biased due to the degree of structural similarity between the target and templates, between all templates used, as well as the refinement performed by the program (A. Fiser, personal communication). Instead, the models were subjected to a more basic quality analysis. *V. vulnificus* models were initially ranked based on the objective function returned by MODELLER, after which the Ramachandran space for each model was analyzed using PROCHECK-NMR [150, 153]. Finally, the models were visually inspected to verify that the secondary structure of each model was consistent with the expected response regulator secondary structure arrangement, and that conserved residues were present in their expected positions.

Model visualization

Models were visualized using PYMOL [154]. Hydrophobic surface plots were produced by highlighting the side-chain atoms of the following residues: I, L, V, G, A, F, C, M, S, T, W, Y, P, and H. In order to visualize subtle differences in side-chain hydrophobicity, we developed a color-coded hydrophobic scale that highlighted side-chains based on a relative hydrophobicity scale derived from averaged physicochemical properties of the amino acid side-chains [155, 156]. Hydrophobic rankings were assigned and color-coded as a gradient, from high to low as follows: ILV (red), GAF (orange), CM (yellow), ST (green) and WYPH (blue). This is shown below in Figure 3.3.

Figure 3.3. Hydrophobic scale used to visualize changes in side chain hydrophobicity. Numerical values are applied to each color code from 5 (red) to 1 (blue).

Results

Modeled Receiver Domains of RRs from V. vulnificus

Of the ninety-two domains predicted to be CheY-like by the SUPERFAMILY database, eighty-one were successfully modeled. A summary of the quality analysis of the final models are shown below in Table 3.1. Following the procedure created to sub-classify the OmpR subfamily of response regulators in *B. subtilis* and *E. coli*, a 15 Å × 15 Å square was centered about the α-helix 1/α-helix 5 interface, and subsequently divided into three vertical strips, each approximately 5 Å wide. This division roughly corresponds to α-helix 5 and the β4-α4 loop (strip 1), the α-helix 1/α-helix 5 interface (strip 2), and to α-helix 1 (strip 3). The average hydrophobic content based on amino acid composition was calculated for each strip using the gradient scale shown above in Figure 3.3, which was then combined with the hydrophobic content totals for the 225 $Å^2$ region in the initial step in model classification. Further sub-classification was conducted through the visual inspection of the model surfaces, studying the amino acid compositions and their positions within each strip, and by the clustering of models given commonalities of hydrophobic patch characteristics. This visual

inspection allows for models that might share similar hydrophobic scores to be placed in different classes, given different amino acid content and positioning.

Figures 3.4 – 3.8 below illustrate the modeled *V. vulnificus* response regulator receiver domains divided into sub-classes based on commonalities of hydrophobic content in the α-helix 1 and α-helix 1/α-helix 5 interface surface, known to possess approximately 75% of all response regulator receiver domain contacts to histidine kinase. Table 3.2 shows a breakdown of the sub-classes developed in this study, including the hydrophobic content scores used to describe each sub-class.

Table 3.1. Model statistics for *V. vulnificus* response regulator receiver domain models. *a* Molecular probability density function as returned by MODELLER, *b* Ramachandran values returned by PROCHECK (M most favored, A additionally allowed, G generously allowed, and D disallowed)

GenBank Accession ID	Sequence	Model	PDF[a]	ϕ ψ(M)[b]	ϕ ψ (A)[b]	ϕ ψ (G)[b]	ϕ ψ (D)[b]
37675668	510-633	B99990003	7048.4771	93.7	2.7	0.9	2.7
37675679	1-124	B99990010	6867.4321	90.2	7.1	1.8	0.9
37675773	1-124	B99990002	6393.1304	94.5	3.7	0.0	1.8
37675909	1-124	B99990004	5862.5029	98.2	1.8	0.0	0.0
37675985	1-124	B99990009	6773.7007	97.4	2.6	0.0	0.0
37675987	1-124	B99990002	6368.793	93.9	5.2	0.0	0.9
37675988	15-128	B99990005	7130.0068	87.7	9.6	1.8	0.9
37675989	1017-1140	B99990004	6436.7632	94.5	5.5	0.0	0.0
37676089	170-299	B99990001	8251.9062	92.2	5.2	1.7	0.9
37676235	189-327	B99990007	7570.7642	90.6	7.9	0.0	1.6
37676244	694-816	B99990007	8573.0625	91.5	5.1	2.5	0.8
37676244	541-660	B99990001	5994.187	93.7	4.5	0.9	0.0
37676245	1-124	B99990002	6536.6138	90.8	6.4	0.9	1.8
37676273	1-124	B99990005	6687.2524	93.8	5.7	0.0	0.9
37676294	727-850	B99990001	7288.2202	93.8	6.2	0.0	0.0
37676307	988-1111	B99990004	7123.8154	92.7	5.5	0.9	0.9
37676308	1-124	B99990005	7174.8574	86.1	7.8	4.3	1.7
37676309	13-126	B99990009	7034.1025	88.3	9.9	0.0	1.8
37676377	11-134	B99990001	7032.3384	91.7	4.6	1.9	1.9
37676469	13-136	B99990002	6494.1978	92.7	7.3	0.0	0.0
37676477	4-128	B99990001	6722.6973	92.9	6.2	0.9	0.0
37676593	1114-1237	B99990002	6242.9028	95.6	3.5	0.9	0.0
37676593	991-1113	B99990004	7318.3735	89.2	8.1	0.9	1.8
37676593	1259-1381	B99990001	6418.7744	95.3	1.9	0.9	1.9
37676596	5-133	B99990002	6515.7104	93.2	6.0	0.0	0.9
37676721	4-127	B99990008	6591.5752	92.5	6.5	0.9	0.0
37676845	1-120	B99990009	6611.8101	93.3	6.7	0.0	0.0
37676847	1-124	B99990004	7049.6147	91.9	8.1	0.0	0.0
37676880	731-850	B99990004	6929.0264	88.1	9.2	1.8	0.9
37676889	754-877	B99990003	6753.6572	92.7	7.3	0.0	0.0
37676982	1-115	B99990002	6453.3516	92.3	5.8	1.9	0.0
37677080	589-712	B99990002	6807.4106	93.0	5.3	0.9	0.9
37677132	987-1110	B99990001	6950.084	93.8	4.5	0.9	0.9
37677133	5-131	B99990010	6244.3418	93.0	5.3	1.8	0.0
37677138	50-172	B99990003	6315.9204	93.4	2.8	2.8	0.9
37677144	1-120	B99990010	6671.5278	91.6	5.6	0.0	2.8
37677310	1-124	B99990006	6552.2593	92.8	5.4	0.9	0.9
37677342	1-124	B99990005	7372.353	91.1	7.1	0.9	0.9
37677345	2-125	B99990006	7217.5562	94.4	4.6	0.9	0.0
37677352	1-119	B99990001	6211.9995	96.2	2.9	0.0	1.0
37678240	1-124	B99990002	6784.4287	91.1	6.2	0.9	1.8
37678325	466-585	B99990005	7275.2979	93.5	5.6	0.0	0.9
37678380	1-120	B99990004	6791.8037	94.4	3.7	0.9	0.9

Table 3.1. continued

37678420	1-124	B99990009	6433.499	95.4	3.7	0.9	0.0
37678828	523-646	B99990006	6651.7944	91.2	8.0	0.9	0.0
37678830	1-124	B99990010	6319.8843	92.8	7.2	0.0	0.0
37678875	1-124	B99990004	6933.8008	92.7	4.5	0.9	1.8
37678907	1-120	B99990001	6844.0986	98.1	1.9	0.0	0.0
37679142	178-301	B99990004	7689.5659	94.6	3.6	1.8	0.0
37679248	1-121	B99990009	6697.2803	91.7	7.3	0.0	0.9
37679274	1-124	B99990004	6752.9771	94.5	4.5	0.9	0.0
37679284	1-123	B99990006	6689.3696	94.5	5.5	0.0	0.0
37679341	513-636	B99990007	6890.4097	92.9	6.2	0.9	0.0
37679342	1-124	B99990006	6395.8979	94.6	3.6	0.9	0.9
37679379	13-127	B99990004	6614.4258	92.0	6.2	0.9	0.9
37679417	1-124	B99990009	6716.0537	95.2	3.8	1.0	0.0
37679475	181-309	B99990010	7150.1436	93.2	5.1	1.7	0.0
37679606	1-120	B99990002	6506.1396	94.3	4.8	0.0	1.0
37679788	1-124	B99990008	7325.5444	91.9	6.3	1.8	0.0
37679811	1-124	B99990005	6292.3989	94.7	4.4	0.0	0.9
37679812	486-609	B99990007	6503.3594	92.7	6.4	0.9	0.0
37679852	446-569	B99990009	6162.2197	93.0	5.3	0.9	0.9
37679896	29-153	B99990003	7026.6333	94.7	5.3	0.0	0.0
37679910	22-145	B99990006	6748.2002	92.0	6.2	1.8	0.0
37680253	17-139	B99990002	6903.981	94.6	4.5	0.0	0.9
37680265	449-571	B99990010	7063.2241	92.9	5.4	1.8	0.9
37680289	1-124	B99990009	7045.4712	92.8	5.4	0.9	0.9
37680408	1-124	B99990005	6748.353	92.5	5.6	0.9	0.9
37680448	597-720	B99990003	6859.9502	96.4	3.6	0.0	0.0
37680487	1-124	B99990004	5693.0303	97.3	2.7	0.0	0.0
37680589	31-154	B99990009	6498.6157	95.6	4.4	0.0	0.0
37680601	19-142	B99990010	6381.3999	90.8	8.3	0.0	0.9
37680644	1-124	B99990002	7069.6802	87.6	10.6	1.8	0.0
37680647	5-130	B99990001	6257.0381	93.8	4.5	1.8	0.0
37680667	1-124	B99990003	6624.8872	93.6	5.5	0.0	0.9
37680766	133-256	B99990007	6634.917	97.3	1.8	0.0	0.9
37680766	12-133	B99990009	6876.7959	90.4	7.9	1.8	0.0
37680917	721-844	B99990009	6105.0405	94.4	3.7	1.9	0.0
37681007	661-784	B99990004	6773.022	88.6	9.6	0.9	0.9
37681297	1-124	B99990010	6491.5547	93.7	3.6	0.9	1.8
37681313	1036-1156	B99990003	8959.8457	87.6	10.5	0.0	1.9

Table 3.2. Sub-classification of response regulator receiver domains from *V. vulnificus*. *a* Average strip position hydrophobic content, *b* total hydrophobic content of the α-helix 1/α-helix 5 surface

GenBank Accession ID	Sequence	Sub-class	1[a]	2[a]	3[a]	R[b]	O[b]	Y[b]	G[b]	B[b]
37675987	1-124	1	2.0	3.5	4.7	4	1	1	1	1
37676307	988-1111	1	1.5	5.0	3.7	4	0	0	1	2
37676847	1-124	1	1.0	5.0	4.0	3	0	1	0	1
37678875	1-124	1	1.0	5.0	4.3	4	2	0	0	1
37679852	446-569	1	1.0	5.0	3.3	5	0	1	1	1
37676294	727-850	1	0.0	3.3	4.0	3	2	0	1	1
37675985	1-124	1	1.7	3.8	3.5	3	1	0	3	2
37676244	694-816	1	1.0	5.0	5.0	4	0	0	0	2
37675988	15-128	2	4.5	3.7	3.0	4	1	0	0	2
37677080	589-712	2	3.3	4.3	4.0	5	3	0	0	2
37680448	597-720	2	3.3	4.7	3.0	4	2	0	2	1
37677345	2-125	2	4.5	4.7	3.3	5	2	0	1	1
37677132	987-1110	2	3.0	5.0	4.5	5	1	0	0	1
37678325	466-585	2	3.0	5.0	5.0	5	0	0	0	1
37679417	1-124	2	3.0	5.0	5.0	6	2	0	0	1
37675989	1017-1140	3	3.3	3.0	4.0	3	1	1	0	2
37676593	1114-1237	3	5.0	4.0	5.0	4	0	0	1	0
37676469	13-136	3	3.0	4.0	5.0	3	4	0	1	1
37676880	731-850	4	1.0	4.7	3.0	2	2	0	1	1
37681007	661-784	4	1.0	5.0	4.0	3	2	0	0	1
37680917	721-844	4	1.0	4.3	3.5	2	2	0	1	1
37679896	29-153	4	1.0	4.2	4.3	3	4	1	0	1
37680253	17-139	4	1.0	5.0	4.7	4	1	0	0	2
37680589	31-154	4	1.0	4.7	5.0	3	1	0	0	1
37678380	1-120	5	1.0	4.8	2.0	3	1	0	1	1
37681297	1-124	5	1.5	4.8	0.0	3	1	0	1	1
37679342	1-124	5	1.0	4.7	0.0	2	1	0	0	1
37680647	5-130	5	2.3	4.3	2.0	2	1	1	2	1
37676377	11-134	6	1.3	3.3	1.0	2	0	0	2	4
37676245	1-124	6	1.3	3.0	2.0	1	1	0	4	3
37679274	1-124	6	1.0	3.2	0.0	2	1	0	0	3
37675679	1-124	7	2.7	3.0	3.0	3	0	1	1	3
37676982	1-115	7	1.5	2.8	4.5	2	2	0	1	3
37677138	50-172	7	1.0	3.5	5.0	2	2	0	0	2
37678420	1-124	7	1.0	4.0	4.0	3	1	0	0	3
37679811	1-124	7	2.5	3.4	3.0	4	1	0	2	3
37680601	19-142	7	2.5	3.0	4.5	2	3	0	1	2
37680766	133-256	7	1.0	3.2	5.0	3	1	0	0	3
37676273	1-124	7	1.0	3.8	4.5	3	2	0	0	3
37680487	1-124	7	1.0	3.0	4.0	3	2	0	2	2
37677144	1-120	8	3.3	3.2	0.0	2	3	0	1	2

Table 3.2. continued

37676309	13-126	9	1.0	3.3	4.0	3	1	1	2	1
37676477	4-128	9	3.0	4.0	4.0	2	1	1	0	1
37676593	991-1113	9	1.0	4.0	3.0	1	3	1	1	1
37680408	1-124	9	2.5	3.8	3.0	2	2	1	2	1
37675668	510-633	9	1.0	4.3	5.0	4	0	1	0	1
37679812	486-609	9	1.5	3.3	3.7	2	2	2	1	1
37680265	449-571	9	2.3	4.0	4.0	2	1	2	1	1
37681313	1036-1156	9	1.0	3.6	5.0	3	1	1	0	1
37675909	1-124	10	1.0	3.0	0.0	1	0	1	0	2
37676244	541-660	10	2.5	3.8	1.0	1	2	2	0	2
37679248	1-121	10	2.5	3.0	0.0	1	2	1	1	2
37679606	1-120	10	3.0	4.0	0.0	2	1	1	0	1
37677352	1-119	10	1.0	4.0	1.7	1	1	1	2	2
37676593	1259-1381	11	1.5	3.0	3.8	2	1	2	4	1
37676721	4-127	11	1.5	4.0	5.0	3	1	0	1	1
37679475	181-309	11	3.0	4.0	3.0	1	2	0	1	1
37680289	1-124	11	3.0	4.7	4.0	3	2	1	1	0
37680644	1-124	11	0.0	4.7	4.0	2	2	0	0	0
37677133	5-131	12	2.3	3.3	4.0	1	4	0	2	2
37678240	1-124	12	1.0	3.8	3.0	1	3	0	2	1
37679379	13-127	12	1.0	3.4	4.0	1	4	0	2	1
37680766	12-133	12	3.3	2.5	4.0	1	2	1	1	1
37676889	754-877	13	2.3	4.5	3.0	1	2	1	1	1
37678907	1-120	13	1.3	4.3	3.0	3	1	1	1	3
37677310	1-124	14	1.5	3.8	1.0	1	2	0	3	4
37679284	1-123	14	2.0	3.7	0.0	1	2	0	2	2
37679910	22-145	14	1.5	3.3	4.5	2	3	1	3	1
37680667	1-124	14	3.0	3.2	4.0	1	5	0	1	2
37675773	1-124	15	3.0	3.8	3.5	3	2	1	2	1
37678828	523-646	15	2.5	5.0	3.0	4	1	0	2	1
37679341	513-636	15	2.0	5.0	2.0	3	1	0	2	2
37676235	189-327	15	3.0	4.0	3.0	3	2	0	2	1
37678830	1-124	15	1.0	3.7	3.5	2	1	0	2	2
37676308	1-124	15	1.5	3.7	4.0	3	1	0	3	1
37676845	1-120	16	2.5	3.0	3.5	1	2	0	2	1
37677342	1-124	16	1.0	4.0	2.0	1	0	2	1	2
37679142	178-301	16	4.0	4.0	5.0	1	3	0	0	0
37676089	170-299	16	4.0	4.0	3.5	1	6	0	0	0
37676596	5-133	16	1.0	3.4	4.0	2	1	1	2	1

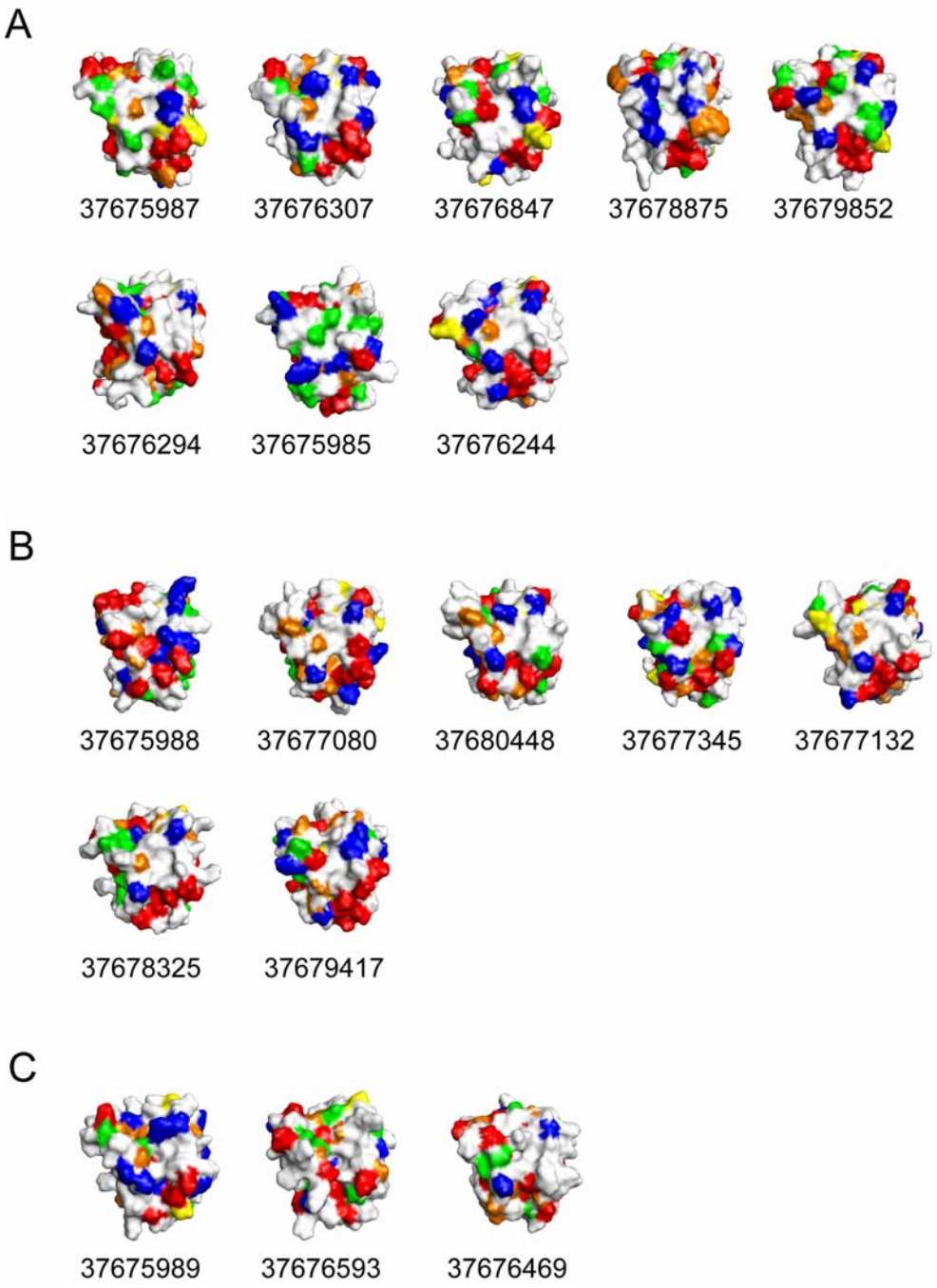


Figure 3.4. Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 1 – 3. *A* Group 1, *B* Group 2, *C* Group 3. The orientation of each molecule is such that α-helix 1 and the α-helix1/α-helix 5 interface are visible.

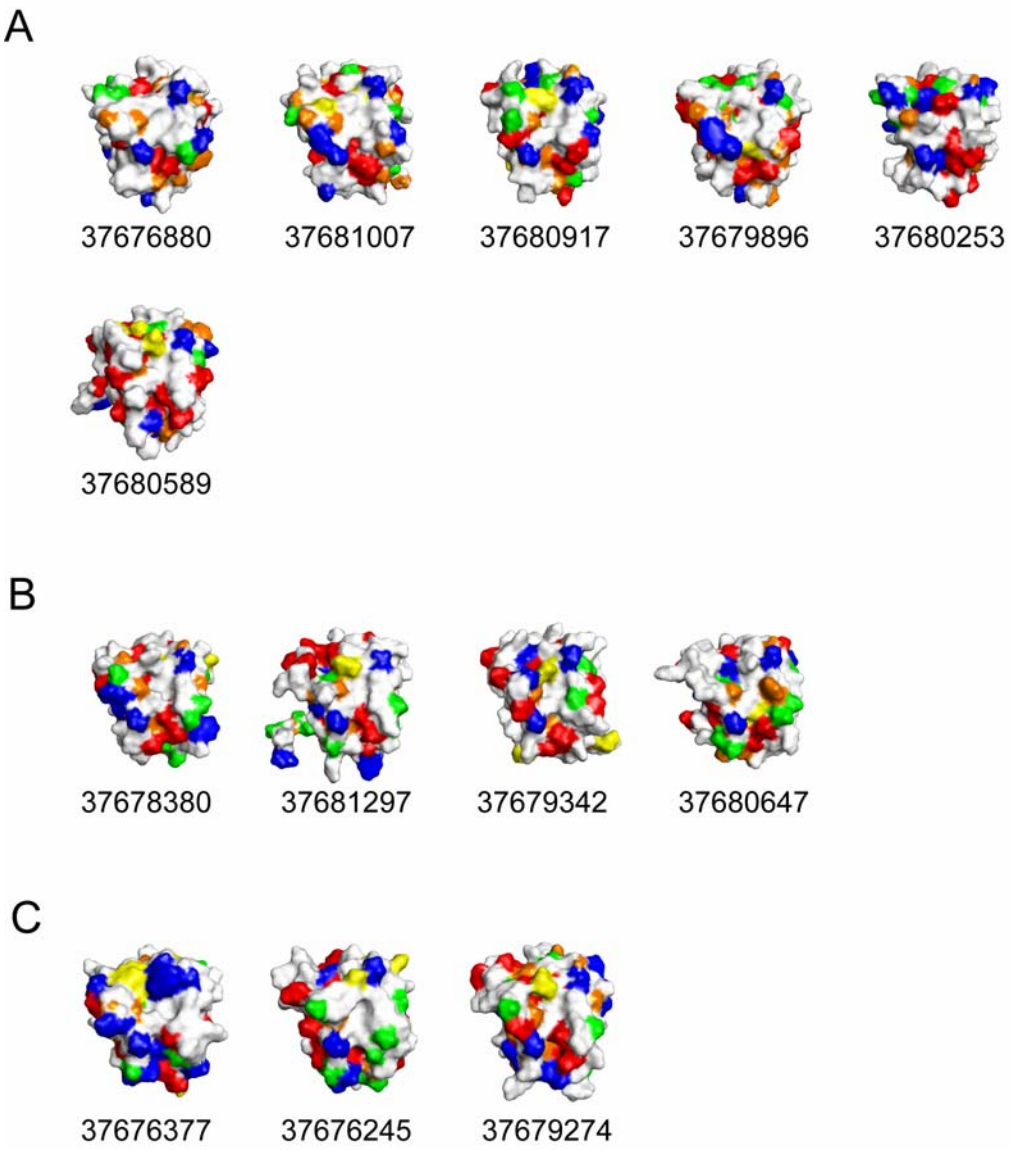


Figure 3.5. Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 4 – 6. *A* Group 4, *B* Group 5, *C* Group 6. The orientation of each molecule is such that α-helix 1 and the α-helix1/α helix-5 interface are visible.

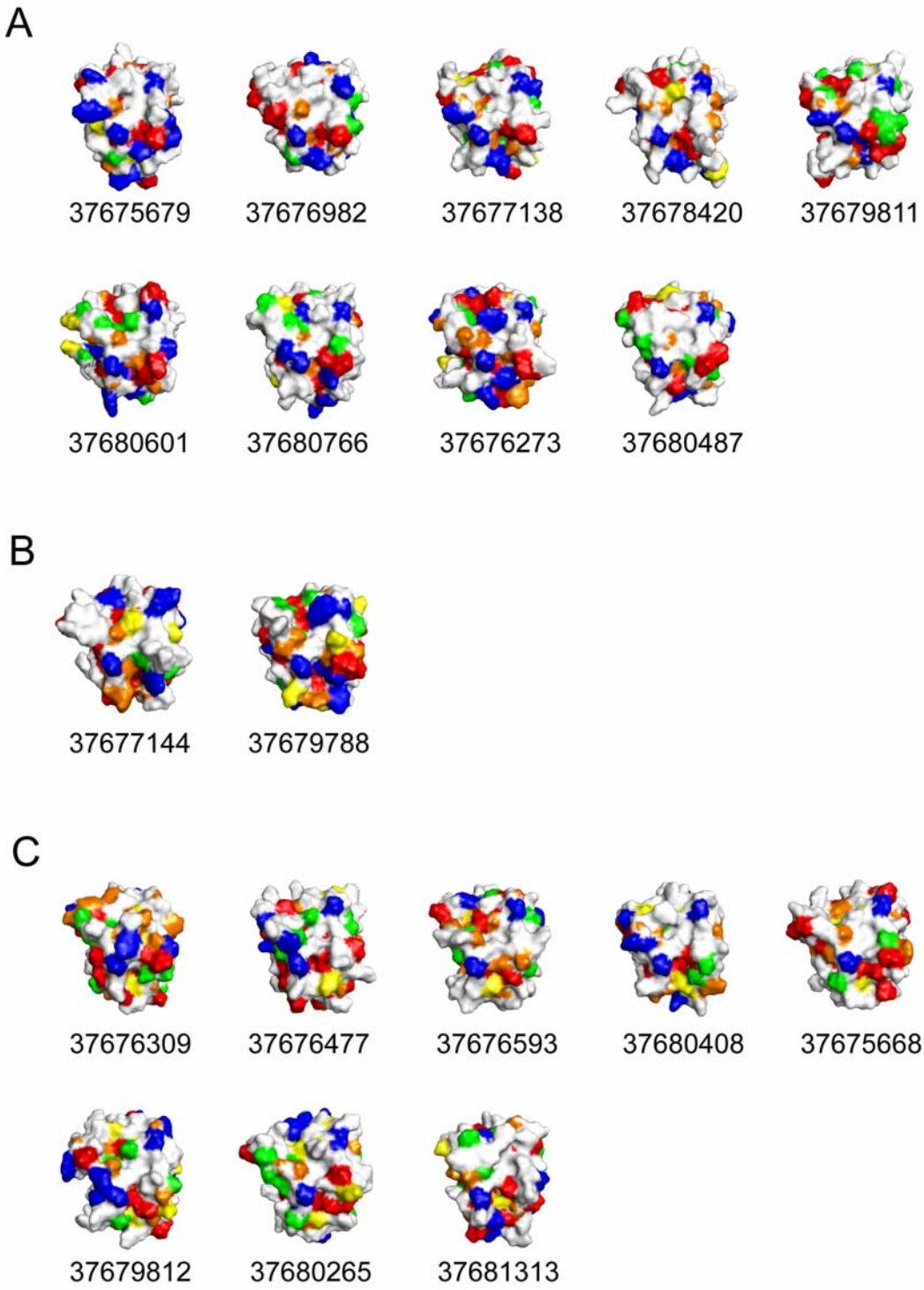


Figure 3.6. Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 7 – 9. *A* Group 7, *B* Group 8, *C* Group 9. The orientation of each molecule is such that α-helix 1 and the α-helix1/α helix-5 interface are visible.

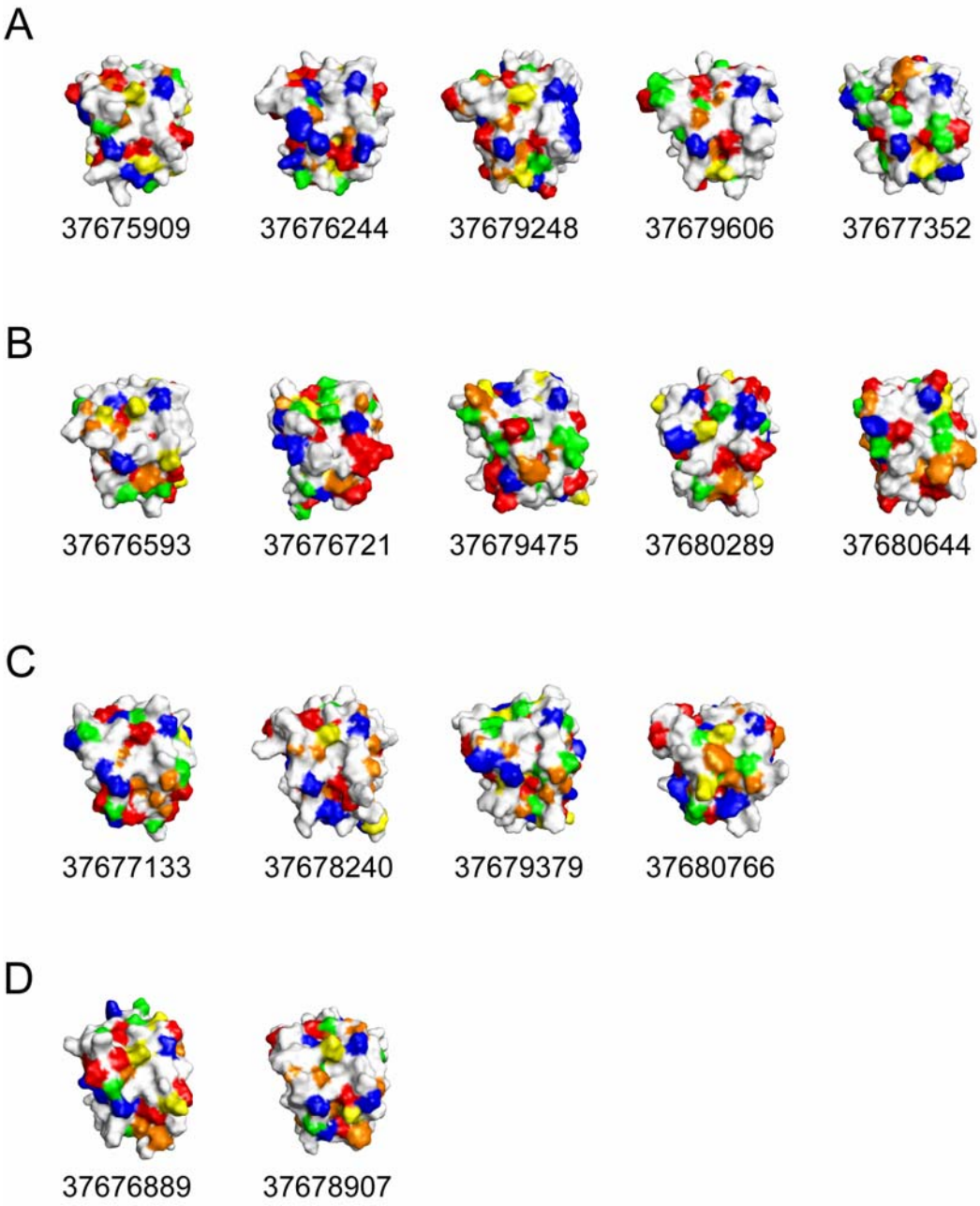


Figure 3.7. Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 10 – 13. *A* Group 10, *B* Group 11, *C* Group 12, *D* Group 13. The orientation of each molecule is such that α-helix 1 and the α-helix1/α helix-5 interface are visible.

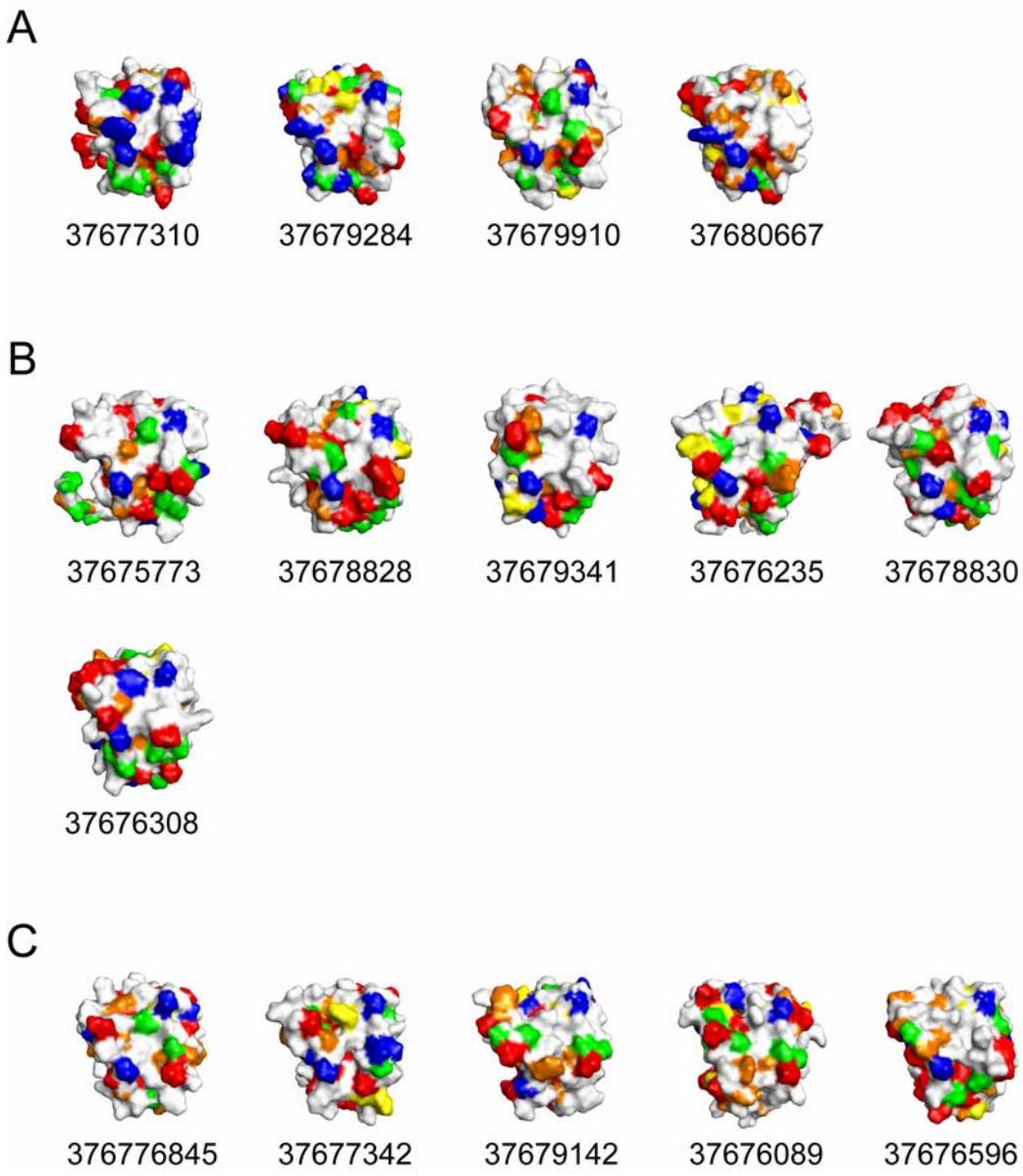


Figure 3.8. Color gradient hydrophobic surface characteristics of *V. vulnificus* models corresponding to Groups 14 – 16. *A* Group 14, *B* Group 15, *C* Group 16. The orientation of each molecule is such that α-helix 1 and the α-helix1/α helix-5 interface are visible.

Discussion

In the prior study by Kojetin et. al., three-dimensional comparative models of the receiver domains of the OmpR sub-family of response regulators from *B. subtilis* and *E. coli* were used to determine if further insight into the relatedness could be achieved among this closely related group of proteins [143]. Prior to this, classification of the proteins comprising the two-component regulatory system had been primarily based on sequence analysis of this histidine kinase and the output domain and activity of the response regulator [49, 137]. Analysis of the surface characteristics of three-dimensional comparative models revealed that sub-classification according to the hydrophobic nature of the response regulator receiver domain – HisKA/Hpt domain interaction surface was plausible and correlated well with previous sequence based classification [143]. In this study we extended the methodologies developed for comparative modeling and classification of *B. subtilis* and *E. coli* OmpR receiver domains to all predicted response regulator receiver domain sequences in the recently sequenced genome of the organism *V. vulnificus* YJ016, in which previous classification had not been conducted. As noted in the previous study of OmpR receiver domains, the *Vibrio* comparative models retain a conserved hydrophobic patch in the HisKA/Hpt domain interaction interface, in which there is considerable variation in the nature of hydrophobic residues comprising this region. The variability in this interface region again provides a method for classification, with implications for the process of discrimination between different HisKA/Hpt domains and is likely the primary determinant in the recognition process.

Arising from the fact that more than three times as many receiver domain sequences were candidates for comparative modeling in this study when compared to the previous work with the OmpR sub-family, several issues arose concerning the scale of this project. Though a great deal of the initial modeling process was semi-automated using scripts written in-house (as mentioned in Methods) a great deal of manual intervention was necessary to derive proper sequence lengths from the annotations in the SUPERFAMILY database to use as targets in the comparative modeling process. This greatly hampered attempts at a high-throughput style modeling procedure. Due simply to the scale of this analysis, another deficiency in this approach becomes apparent: the initial model analysis, based on average hydrophobic strip content coupled with overall hydrophobic content, lacks detailed positional information. For example, an average hydrophobic content of "1" in strip 1 is not indicative of the relative positions of the one or more WYPH residues that would be present. This shortcoming, of course, is remedied by visual inspection of the groupings assembled by strip scores and hydrophobic content alone. However, as the number of models increases, the amount of effort to visually analyze them becomes increasingly laborious. This suggests that further steps might be employed to good effect in the initial characterization prior to visual inspection. An additional step to consider is the use of another set of three 5 Å wide strips, which would be overlaid horizontally atop the initial three vertical strips. This 'six strip analysis' would then be capable of providing a pseudo-Cartesian two-coordinate system which would then better define the positional information of hydrophobic residues.

In addition to a 'six strip analysis' another possible means to better quantify hydrophobicity

and simplify the process of sub-classification would be to incorporate and use a weighting factor in the hydrophobic score of a given residue. This weighting factor could be taken from the relative solvent accessible surface area of each residue, as is possible using the program NACCESS [108]. Residues whose side-chains are entirely solvent accessible would retain their full hydrophobic score, while residues whose side-chains are only partially solvent exposed would have their hydrophobic value scaled according to the percentage solvent exposed. Addition of this weighting factor into hydrophobic score calculation would avoid overly biasing strip scores by residues that are minimally solvent exposed.

Future work using the methodology that would truly complement both this study and the study of Kojetin, et. al. would be homology modeling based sub-classification of the histidine phosphotransferase (HPt) four-helix bundle domain of the sensor histidine kinase. There would be several very interesting aspects to such a study. First, it would answer several questions concerning this methodology, most notably whether or not sub-classification via surface hydrophobicity is able to tease out sub-families of histidine sensory kinases similar to the sequence-based analyses as was seen our previous study of response regulator receiver domains. Secondly, it would be rather interesting to see how well the response regulator interaction surface of the four-helix bundle Hpt domain matches with the surface of cognate response regulators. Though the above analysis would be rather large in scale, it would have several benefits, not the least of which would be identifying possible crosstalk between two component signaling pathways.

Chapter 4: Conformational Changes in the Response Regulator Spo0F Induced by Oxidative Damage

The work in this chapter was performed in collaboration with Dr. Joshua Sharp and Dr. Kenneth Tomer of The National Institute of Environmental Health Sciences.

Abstract

When most proteins undergo oxidative damage, they yield a variety of products containing oxidative damage at many sites, most of which are modified substoichiometrically. The resulting complex mixture of products is not amenable to high-resolution structural analyses. Previous methods of structural analysis have relied upon either generalized structural analyses such as circular dichroism, or upon the creation of a battery of mutants to try to isolate single residue damage effects. Presented here is a methodology using mass spectrometry to measure the kinetics of oxidation at many sites simultaneously. Previous studies have shown that these kinetics are determined by the chemical nature of the oxidation site and by the accessibility of that site to the radical. By measuring deviations in the rate of oxidation from the expected pseudo-first order kinetics, it is possible to detect and characterize local structural changes due to the oxidative damage. The application of this new technique is demonstrated on the protein Spo0F, a regulator of sporulation in *Bacillus subtilis*. Circular dichroism studies suggest a partial loss of helical structure of Spo0F as a result of oxidative damage. The mass spectrometric approach shows that oxidation causes a

three-stage conformational change in Spo0F. Furthermore, the most dramatic structural changes observed affect only the region surrounding the active site, while the remainder of the structure remains relatively unperturbed. Finally, the specific oxidation event initiating the conformational change at the active site of Spo0F occurs at Met81, a partially-conserved methionine in the CheY superfamily.

Introduction

Oxidative stress is a cellular state defined as an imbalance in the distribution of intracellular reactive oxygen species (prooxidants) and cellular reductive agents (antioxidants) in the favor of prooxidant factors. Cumulatively, oxidative stress comprises any conditions which favor the accumulation of reactive oxygen species as well as any conditions leading to the depletion of antioxidant molecules. Such circumstances can arise from metabolic processes - which are the main sources of endogenously produced superoxide anions and hydrogen peroxide molecules - and exposure to metals, ionizing radiation, and redox-active chemicals – which can lead to the formation of other reactive oxygen species, such as the hydroxyl radical [3]. The condition of oxidative stress is of great biological relevance, primarily because reactive oxygen species are highly deleterious to cellular survival as they are damaging to all cellular components - protein, DNA, RNA and lipid. It is generally accepted that challenge by a number of stresses, including oxidative stress, decreases the proportion of cells in a culture undergoing sporulation. Recent work indicates that shear stress leads to 20-fold reduction in the ability of cells to sporulate due to the generation of reactive oxygen

species (ROS), specifically superoxide radical derivatives, within the cell. It was further found that induction of oxidative stress through the addition of exogenous hypochlorus acid or hydrogen peroxide also caused a similar inhibition of sporulation; reduction of the level of intracellular ROS rescues sporulation [157].

A particularly appealing target for the down-regulation of sporulation is the signal transduction pathway that is responsible for its initiation. The initiation of sporulation in *Bacillus subtilis* is tightly regulated by an exquisite signaling mechanism referred to as the Stage 0 multicomponent phosphorelay pathway [47, 59, 61, 66, 69, 70, 77, 137]. In response to distinct (and as of yet uncharacterized) environmental signals arising from nutritive deprivation and cell density, the primary sensor kinase KinA autophosphorylates on a conserved histidine. KinA transfers this phosphoryl group to a conserved Asp residue in the conserved Asp pocket of the single domain response regulator protein, Spo0F. The phosphoryl group is subsequently transferred from Spo0F to a conserved histidine on the phosphotransferase, Spo0B, and ultimately to the response regulator domain of the 'master switch' of sporulation, Spo0A. The phosphorylation of Spo0A serves to activate the protein. Activated Spo0A-P is a transcription factor, initiating the cascade of gene activation required for the proper formation of a viable spore [3]. Considering that oxidative stress leads to a downregulation of sporulation, it is reasonable to presume that a mechanism may exist by which intracellular reactive oxygen species act upon components of the phosphorelay pathway in a negative fashion.

This study focuses upon delineating the effects of oxidative damage to the response regulator Spo0F. Spo0F is a 124 residue, α5/β5 single domain protein in which the helices are arranged around a central parallel β-sheet core, helices 1 and 5 on one face of the β-sheet, helices 2, 3, and 4 on the other face. At the 'bottom' of the structure, removed from the phosphorylation site, are short α-β loops connecting elements of secondary structure (e.g. α1-β2, α2-β3 etc.). As with all other response regulator proteins, Spo0F has a conserved aspartic acid pocket at the 'top' of the protein composed of Asp residues 10, 11 and 54, the residue phosphorylated by KinA. This Asp pocket is surrounded by five functionally critical β-α loops (β1-α1 to β5-α5) [136]. Residues making up the β-α loops as well as a subset of helical surfaces are essential for the activity of Spo0F, directing the proper interactions with KinA, the phosphotransferase Spo0B and the phosphatase RapB [139, 158]. Additionally, metal binding studies have indicated a distinct role of the β-α loops in binding a divalent Mg^{2+}, which is required for phosphorylation. Backbone dynamic studies via NMR have revealed the key role of the β-α loop regions in an intramolecular communications network within Spo0F responsible for transmitting the state of phosphorylation throughout the protein [159, 160]. Indeed, all the known molecular recognition and interaction properties of Spo0F are contained within the β-α loops and the N-terminal end of the α1 helix which is contiguous with the β1-α1 loop. Consequently, any impacts in this region would significantly impede function.

Protein oxidative surface mapping has been used previously to describe the topology of a

number of proteins in solution [161-164]. The rate of oxidation has been shown to be largely dependent on a combination of the inherent reactivity of the oxidized sidechain and the solvent accessibility of that sidechain [161-166]. As such, a careful measurement of the rate

of oxidation at a particular site can be used to infer the details of conformational changes altering the solvent accessibility of that site. Presented below is a comprehensive analysis of the structural effects of oxidative damage on the response regulator Spo0F as studied using quadrupole time-of-flight mass spectrometry (Q-TOF MS). These results are then compared with those generated using circular dichroism - a standard analytical technique for studying structural changes due to oxidative damage. Protein oxidative surface mapping via mass spectrometry not only agrees with the data generated by circular dichroism, but is able to provide much more detail about both the nature of the structural changes and the cause of the most dramatic conformational shift detected. These data were generated using only small amounts of protein in one experiment, with no mutagenesis or other modification of the analyte required. This technique therefore represents a significant advance in the methodology for studying structural effects of oxidative damage to proteins.

Our previous NMR structure and dynamics studies show that the β-α loops have some conformational flexibility on the micro- to millisecond timescale [136, 160]. Nevertheless, these loops cluster at the 'top' of the protein. In this study we show conclusively that, in response to mild oxidation, a conformational change at Met81 triggers a localized conformational change that relaxes and opens up the cluster of β-α loops around the

phosphorylation site. This conformational event severely affects the structure of the regions previously identified as necessary for Spo0F function, and therefore would inhibit the organism's ability to initiate sporulation.

Methods

Sample Preparation for Circular Dichroism and Mass Spectrometry

Purified Spo0F was diluted to working concentrations of 10 μM or 20 μM in 2mM air-saturated ammonium bicarbonate, pH 6.8. Bovine catalase was added to a final concentration of 10 nM tetramer to eliminate hydrogen peroxide generated during radiolysis. 100 μl aliquots were placed in 500 μL tubes and exposed to gamma irradiation in a ^{137}Cs dual source gamma irradiator at a dosage rate of 0.0236 kGy/minute for 0, 10, 20, 30, 40, 50, and 60 minutes at 23º C. The samples were then rotated on a turntable at 60 rpm during exposure to ensure equal radiation exposure for all samples. The 20 μM Spo0F samples were used for circular dichroism spectroscopy analysis while the 10 μM Spo0F samples heated to 85 º C for one hour to denature the protein, whereupon 2 μL of resuspended trypsin was added (20 μg of lyophilized sequencing grade modified trypsin dissolved in 40 μL of trypsin resuspension buffer) to each irradiated Spo0F aliquot. The samples were then placed in a rotating incubator at 37º for 48 hours.

Circular Dichroism Spectroscopy

The 20 μM samples were analyzed using a Jasco J-810 circular dichroism spectrophotometer

(Jasco, Easton MD) in a quartz cell with a 1 mm path length. All spectra were obtained in quadruplicate, and the four signals were averaged together for the final output. The CD spectrum of 2 mM ammonium bicarbonate was then subtracted from all Spo0F spectra to eliminate signal from the buffer.

Nanoelectrospray-Quadropol Time of Flight Mass Spectrometry

The 10 μM irradiated Spo0F aliquots were mixed 50:50 with acetonitrile with 0.2% formic acid and injected directly into a hybrid quadrupole-time of flight Micromass Q-TOF Ultima Global mass spectrometer using a Waters nanoflow electrospray source. Signal was accumulated for nine minutes for each sample. These nine minute accumulations were then split into three separate datasets of three minutes accumulation each, and each of the data sets were independently averaged. When there was insufficient signal in the three datasets to generate three values, all three were combined and averaged, and thus a single value was used. Peptides were assigned by high resolution mass measurement and tandem MS/MS analysis when possible. The resulting mass spectra were used to determine the kinetics of oxidation by measuring the signal intensity of the unoxidized species of each peptide and comparing the signal intensity with that of each of the various oxidized species. The average number of oxidations per peptide were determined by the equation:

$$R = ([M+O] + 2[M+2O] + \ldots)/([M] + [M+O] + [M+2O] + \ldots) \quad \text{(Equation 1)}$$

where [M] is the ion abundance of the unmodified peptide, [M+O] is the abundance of the peptide plus 16 Da, [M+2O] is the abundance of the peptide plus 32 Da, etc. The rate of oxidation was determined by dividing the average number of oxidations per peptide by the radiation dosage in kiloGray (kGy). When multiple charge states could be measured, the charge state that showed the lowest standard deviation in the amount of oxidation across all irradiation dosages was used for analysis to minimize experimental error. When multiple overlapping peptides were detected, the peptide with the highest signal level was used for analysis (usually representing the completed digestion product).

Solvent Accessibility Analysis

Solvent accessibility analyses were performed using the GETAREA 1.1 server available at http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl. The minimized average NMR solution structure (PDB ID 2FSP) was as a template. Solvent accessibility was then calculated for each individual atom.

Results

Circular Dichroism Spectroscopy

In order to examine the structural effects of oxidative damage to Spo0F we first subjected all irradiated samples to CD analysis. The CD spectra are shown in Figure 4.1. From analysis of the CD spectra, the protein appears to lose helical character gradually from 0 to 0.472 kGy irradiation, while the local minimum remains at approximately 209 nm. Between 0.472 kGy

and 0.708 kGy exposure, a more dramatic shift occurs, with the local minimum shifting to approximately 207 nm. Relatively little change occurs between 0.708 kGy and 0.944 kGy exposure. Between 0.944 kGy and 1.18 kGy exposure, another large shift in the local minimum occurs, from ~207 nm to ~205 nm. Between the final 1.180 kGy and 1.416 kGy exposures, little change occurs. The neural network program K2d [167] was utilized in an attempt to deconvolute the CD spectra recorded at the varying radiation doses, but the maximum error reported by the program was too high for confidant analysis. Manual interpretation of these CD spectra indicates a general loss of helical structure with a commensurate increase in random coil content, as seen by a blue shift of the minimum at 209 nm and a loss of ellipticity at 195 nm. In contrast to this loss of helical character, the local minimum at 218 nm remains unshifted throughout the experiment, although it does lose ellipticity. The above data indicates that oxidative damage leads to partial conversion of α-helix to random coil, with the largest transitions occurring between 0.472 kGy and 0.708 kGy, and 0.944 kGy and 1.180 kGy exposure. Most interestingly, overlay of the CD spectra resulted in a clear isodichroic point at 206 nm. This suggested a two-state transition between the native conformation and a single primary oxidized conformation. The presence of a single primary oxidized conformation that dominates the contribution of the wide variety of oxidation products suggested that oxidation at a single highly reactive residue in Spo0F lead to a drastic conformational change.

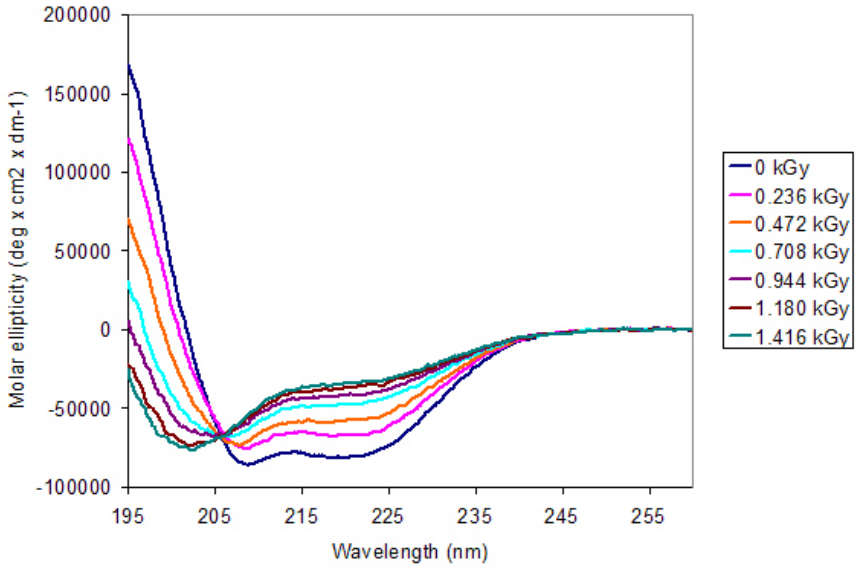


Figure 4.1. CD spectra of Spo0F after varying amounts of γ-ray induced oxidative damage.

Hydroxyl radical generation by low-dose continuous gamma radiolysis results in a continuous, low concentration of hydroxyl radicals, where the actual concentration of radical is based on the rate of hydroxyl radical formation (dependent on the radiation dose rate) minus the rate of hydroxyl radical consumption by various components of the analyte, including the buffer system and consumption of the radical by radical-radical recombination. Though the actual concentration of radical cannot be calculated due to unknown rates of radical reaction with various buffer components, we can state that the radical concentration is constant, and therefore the rate of side chain-radical reaction follows pseudo-first order kinetics, dependent solely upon the concentration of unreacted side chain [168]. Under these conditions the rate of oxidation of an amino acid target will be constant as long as the

concentration of the unoxidized target is much greater than the concentration of hydroxyl radical. This rate of reaction is determined by two factors: i) the inherent chemical reactivity of the amino acid target and ii) the accessibility of the target to the radical [169-172]. Because the chemical reactivity of the target will remain essentially unchanged, by measuring the apparent rate of oxidation and plotting it as a function of radiation dose we can measure changes in solvent accessibility as a function of oxidation. Changes in the apparent rate of oxidation would then indicate a conformational change affecting the solvent accessibility of the oxidation target. In this work, the kinetics of oxidation were provided as oxidation events per molecule per kiloGray of radiation exposure so that future experiments that may be performed using different concentrations of protein and different intensities of radiation may be easily compared.

The rates of oxidation for eleven different regions of Spo0F, containing a minimum of fourteen discreet oxidation sites, were examined simultaneously. 92% sequence coverage was obtained for the protein; the only regions for which no data was collected consisted of N-terminal residues 1 through 5, residues 68 to 70, and C-terminal residues 123 and 124. As the structure of Spo0F had previously been solved by both X-ray crystallography and solution NMR spectroscopy, oxidation rate analysis could be examined in the context of the native structure, and determine which portions of the structure (if any) change conformation as a result of oxidative damage.

Figure 4.2 shows the oxidation kinetics for peptide 6-16. MS/MS analysis (not shown) of

this peptide demonstrated that all detectable oxidation occurred at Tyr13. At the 0.236 kGy dose the amount of oxidation was below the limits of detection. From 0.472 kGy to 0.708 kGy the rate of oxidation showed a slight increase, indicating that Tyr13 experienced a slight increase in solvent accessibility. After 0.708 kGy, the rate remained rather linear throughout the rest of the experiment. In the native structure of Spo0F Tyr13 is almost completely and therefore the kinetics of oxidation at this residue would be insensitive to local conformational change.

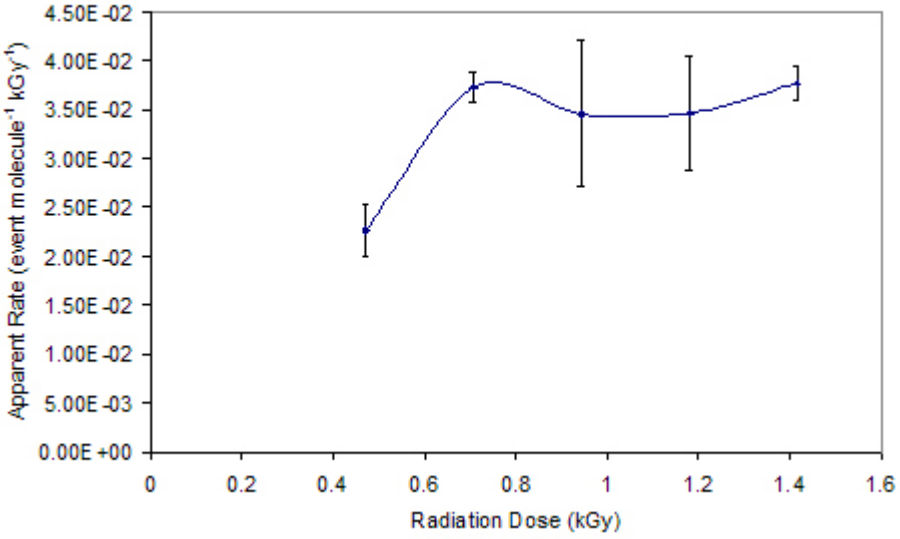


Figure 4.2. The oxidation rate for peptide 6 - 16

The oxidation kinetics for peptide 17-25 are shown in Figure 4.3. MS/MS analysis (not shown) indicated that all detectable oxidation occurred at Leu18. Between 0.236 kGy and

0.472 kGy, an increase in the rate of oxidation indicated a slight increase in solvent accessibility of Leu18 due to conformational change. This was followed by a region of stable rate of oxidation from 0.472 kGy to 0.708 kGy. Between 0.708 kGy and 0.944 kGy, a large increase in the rate of oxidation occurs, indicating a large conformational change leading to a further increase in the solvent accessibility of Leu18. This was then followed by a region of stable oxidation kinetics between 0.944 kGy and 1.180 kGy. Finally, the rate of oxidation began to increase significantly again between 1.180 kGy and 1.416 kGy.

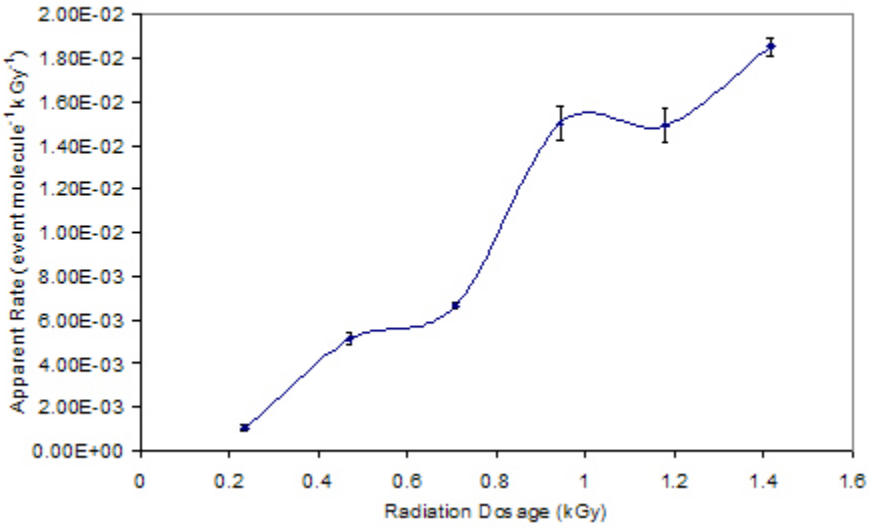


Figure 4.3. The oxidation rate for peptide 17 - 25

Figure 4.4 shows the oxidation kinetics of peptide 26-45. Tandem MS was unable to determine the exact site(s) of oxidation due to insufficient signal. However, examination of the sequence of the peptide in the context of the native structure suggested that oxidation

most likely occurred on Tyr28 and/or Phe31. Insufficient oxidation occurred at 0.236 kGy to quantify the kinetics of oxidation; however, between 0.472 kGy and 0.708 kGy the rate of oxidation remained the same, suggesting a region of conformational stability about the oxidation target(s) in this peptide. Between 0.708 kGy and 0.944 kGy a slight increase was seen in the rate of oxidation. For the remainder of the experiment, the rate of oxidation remained the same, indicating that the structure around the oxidation site(s) retained a stable conformation.

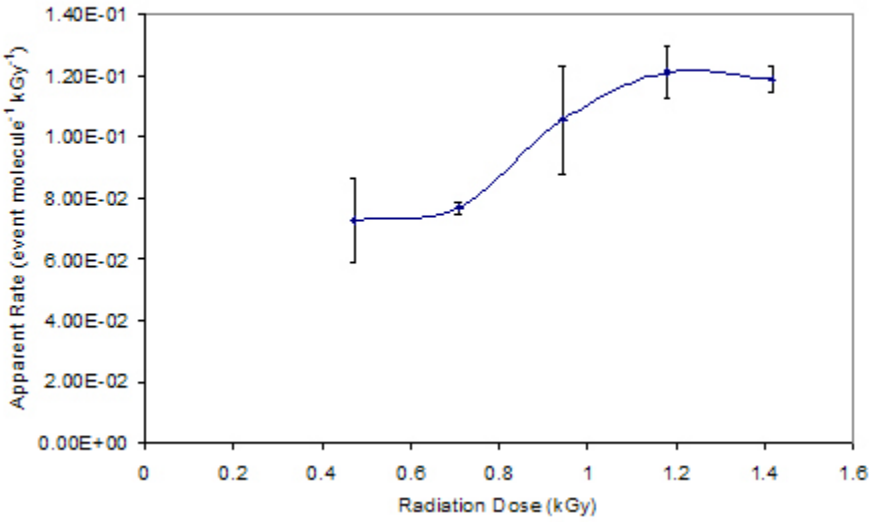


Figure 4.4. The oxidation rate for peptide 26 - 45

The kinetics of oxidation for peptide 46-56 are shown in Figure 4.5. MS/MS analysis determined that all detectable oxidation occurred at Met55 (not shown). This peptide showed a slight increase in the rate of oxidation between 0.236 kGy and 0.472 kGy. Between 0.472

kGy and 0.708 kGy little change in the kinetics of oxidation indicated that there was a region of relative structural stability. From 0.708 kGy to 0.944 kGy, the rate of oxidation increased, indicating an increase in the accessibility of Met55 to the radical. No change in rate was detected from the 0.944 kGy dose to the end of the experiment.

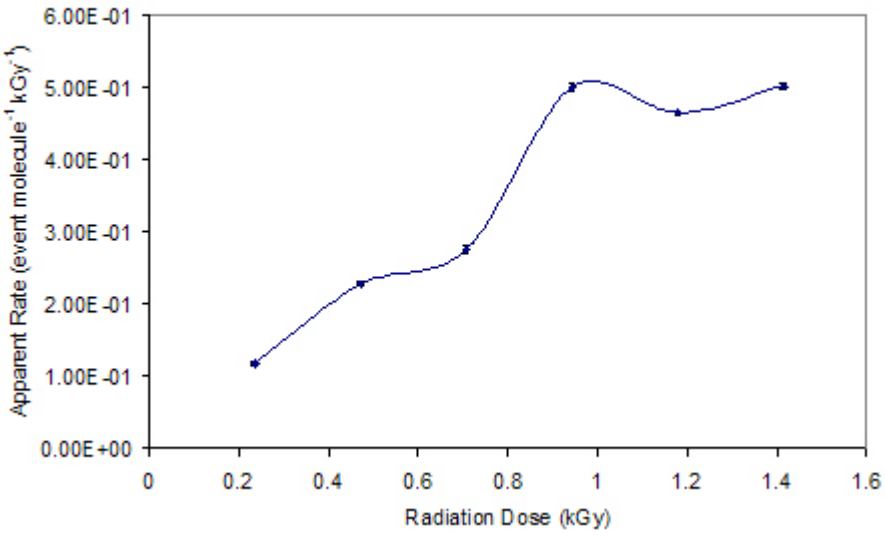


Figure 4.5. The oxidation rate for peptide 46 - 56

The kinetics of oxidation of peptide 57-67 is shown in Figure 4.6. MS/MS analysis indicated that all detectable oxidation occurred at Met60 (not shown). This peptide showed a slight increase in the kinetics of oxidation from 0.236 kGy to 0.472 kGy, followed by a period of conformational stability form 0.472 kGy to 0.708 kGy. From 0.708 kGy to 0.944 kGy, there was a dramatic increase in the rate of oxidation at Met60, indicating a change in

conformation that resulted in a large increase in the solvent accessibility of Met60. From 0.944 kGy to 1.180 kGy, relative structural stability resulted in stable oxidation kinetics. From 1.180 kGy to 1.416 kGy, the kinetics of oxidation showed another large increase, indicating further conformational change affecting the solvent accessibility of Met60.

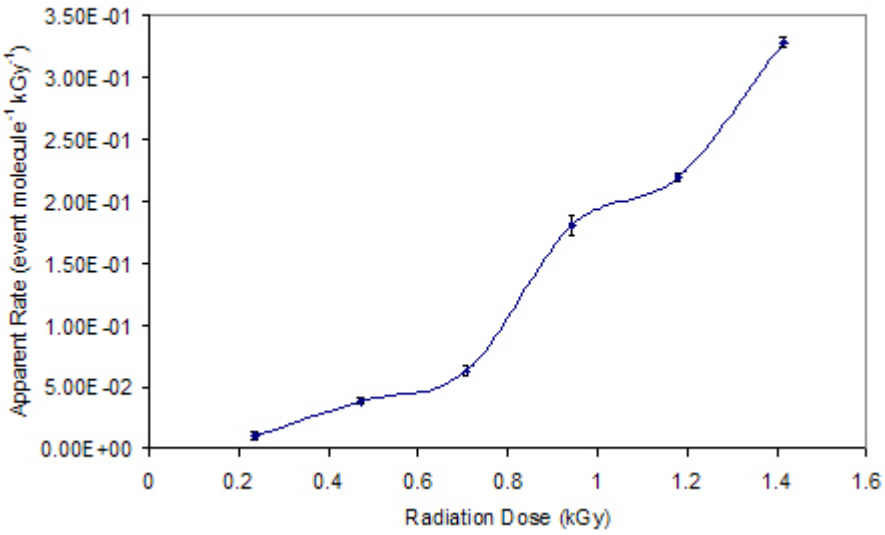


Figure 4.6. The oxidation rate for peptide 57 - 67

The kinetics of oxidation of peptide 71-77 are shown in Figure 4.7. Insufficient signal existed to determine the precise site(s) of oxidation by MS/MS. However, analysis of the primary sequence of the peptide in the context of the native structure suggested that Ile72 was the probable oxidation target. The rate of oxidation of this peptide was very low, below the limit of detection until 0.708 kGy exposure. The kinetics of oxidation for the remainder

of the experiment after 0.708 kGy could be measured, and showed no significant change. All previously measured peptides that demonstrated a large conformational change did so between 0.708 kGy and 0.944 kGy, or between 1.180 kGy and 1.416 kGy. Peptide 71-77 showed significant increases in the rate of oxidation over these dosages. These data suggest that no gross conformational change took place that increased the solvent accessibility of Ile72.

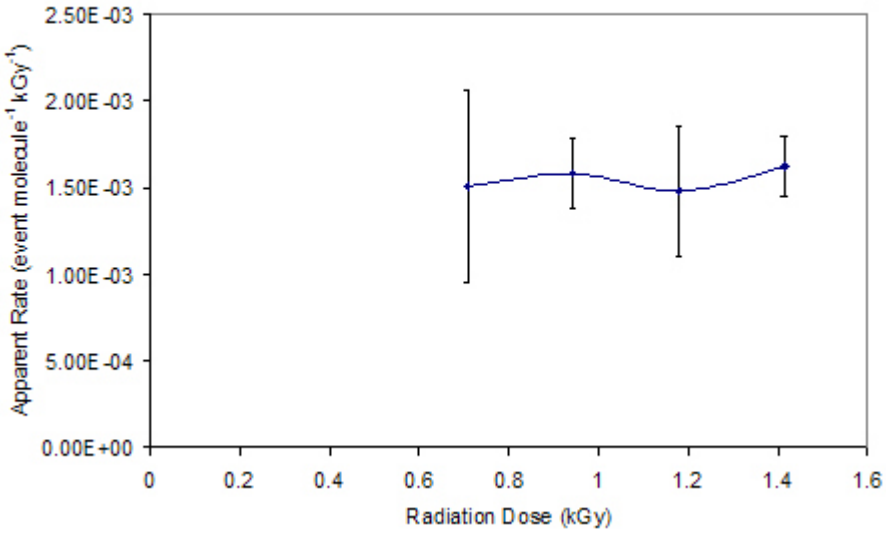


Figure 4.7. The oxidation rate for peptide 71 - 77

The kinetics of oxidation of peptide 78-94 are shown in Figure 4.8. Analysis of the dioxidized peptide (M+32 Da) by tandem MS determined that oxidation occurred at both Met81 and Met89 (not shown). Further MS/MS analysis of the singly oxidized protein also

showed that the oxidation was split between Met81 and Met89. Accurate quantitation of the relative amount of oxidation at each site was not possible from these data due to the effects of methionine oxidation on the fragmentation pattern of the peptide. The oxidation kinetics at the 0.236 kGy dose could not be determined due to very low signal. It is interesting to note that over the course of the experiment the rate of oxidation of this peptide decreased. Analysis of the amount of oxidation suggested that the reaction left the region of linear pseudo-zero order kinetics, following first order kinetics as the oxidation target became saturated. Despite a change in kinetic rate order, conformational changes surrounding these

two regions would be reflected by an increase in kinetics, which was not observed for this peptide. The rate of decrease does slow between 0.708 kGy and 0.944 kGy, which may reflect a modest increase in the solvent accessibility of one or both of the methionines. Given that the sulfur of Met89 is already highly exposed in the native structure (14.69 $Å^2$ of solvent accessible area compared with 0.06 $Å^2$ of solvent accessible area for Met81) full unfolding of the region surrounding Met89 would only be expected to result in a modest increase in the rate of oxidation.

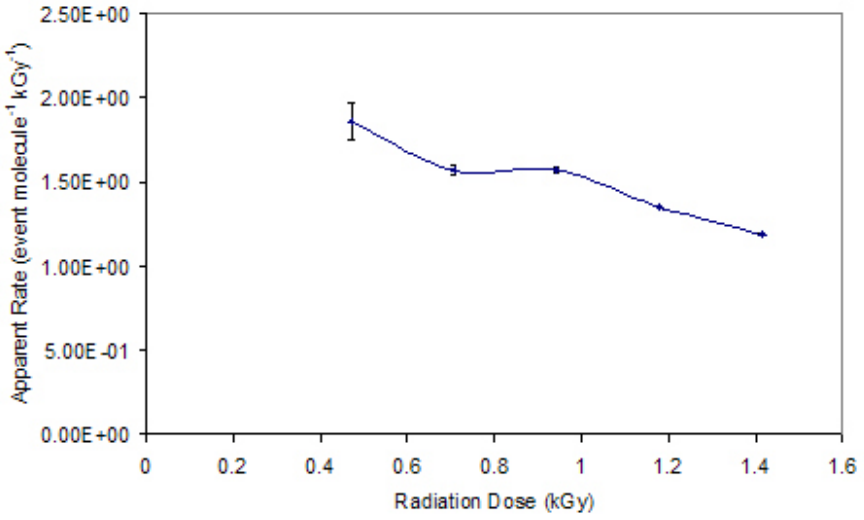


Figure 4.8. The oxidation rate for peptide 78 - 94

Conversely, a conformational change increasing the solvent accessibility of Met81 would be expected to result in a much larger increase in the kinetics of oxidation, which again was not observed. Fortunately, an additional, independent probe for this region was available. Oxidative modification of an acidic residue can lead to oxidative decarboxylation, which results in a net loss of 30 Da. For peptide 78-94, a product showing a loss of 30 Da from the unmodified peptide (as well as all of the peptides oxidized at Met81 and Met88) was detected. Kinetics of oxidative decarboxylation of peptide 78-94 are shown in Figure 4.9. Tandem MS analysis revealed that all detectable oxidative decarboxylation occurred at Asp88 (not shown). The oxidative decarboxylation product was below the limits of detection at 0.236 kGy. Between 0.472 kGy and 0.708 kGy, no increase in the kinetics of oxidation

was detected, indicating a region of conformational stability. Between 0.708 kGy s and 0.944 kGy, a dramatic increase in the rate of oxidation was detected, which pointed to a significant conformational change that increased the solvent accessibility of Asp88. The rate of oxidation continued to increase between 0.944 kGy and 1.180 kGy, although the experimental error (due to the very low signal of these products) made the extent of this increase difficult to accurately determine. However, it can be said that the rate of oxidation did increase significantly between 0.944 kGy and 1.180 kGy, indicating continued conformational change.

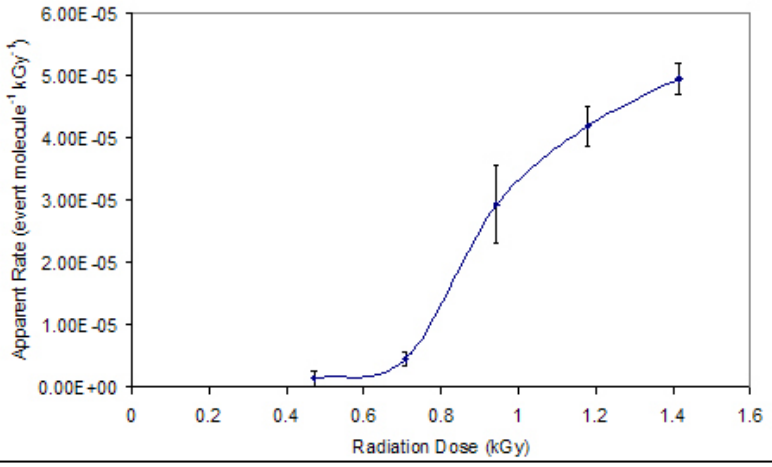


Figure 4.9. The rate of oxidative decarboxylation for peptide 78 - 94

Figure 4.10 shows the kinetics of oxidation for peptide 95-104. Positive identification of the site(s) of oxidation could not be determined from MS/MS, due to insufficient signal and fragmentation in the tandem mass spectra. However, examination of the partial MS/MS spectra combined with the sequence of the peptide in the context of the structure of Spo0F

indicated that the most likely site of oxidation is Phe102. Between 0.236 kGy and 0.472 kGy, there was an increase in the rate of oxidation, followed by a region of structural and rate stability from 0.472 kGy to 0.708 kGy. A dramatic increase was seen from 0.708 kGy to 0.944 kGy, which indicated a large conformational change similar to those detected in several peptides described above. However, unlike the above peptides, the kinetics of oxidation for peptide 95-104 did not have a period of relative stability from 0.944 kGy to 1.180 kGy; rather, the kinetics of oxidation continued to increase, which indicated that the change in solvent accessibility experienced by this particular oxidation target continued to increase throughout the remainder of the reaction, without the stable intermediate seen in other regions.

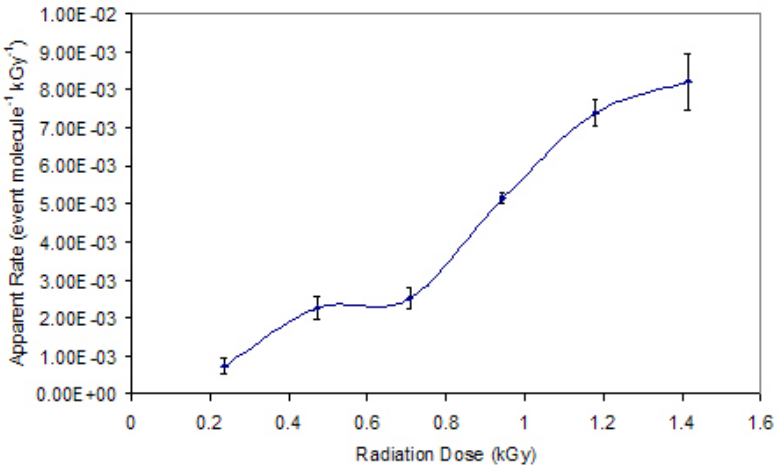


Figure 4.10. The oxidation rate for peptide 95 - 104

The kinetics of oxidation of peptide 105-112 are shown in Figure 4.11. Oxidation of peptide 105-112 occurred very slowly. While the unoxidized peptide could be readily detected at all dosages, the amount of oxidized peptide was below the limits of detection until 0.944 kGy. Signal from the oxidized peptide was inadequate for MS/MS determination of the site(s) of oxidation; analysis of the peptide sequence and the native structure suggested that Phe106, which is almost fully buried in the native structure, was the likely oxidation target. No significant difference in the rate of oxidation could be determined from the 0.944 kGy dose to the end of the reaction at 1.416 kGy. As this residue is almost completely buried, this suggested that this target does not experience any significant conformational change.

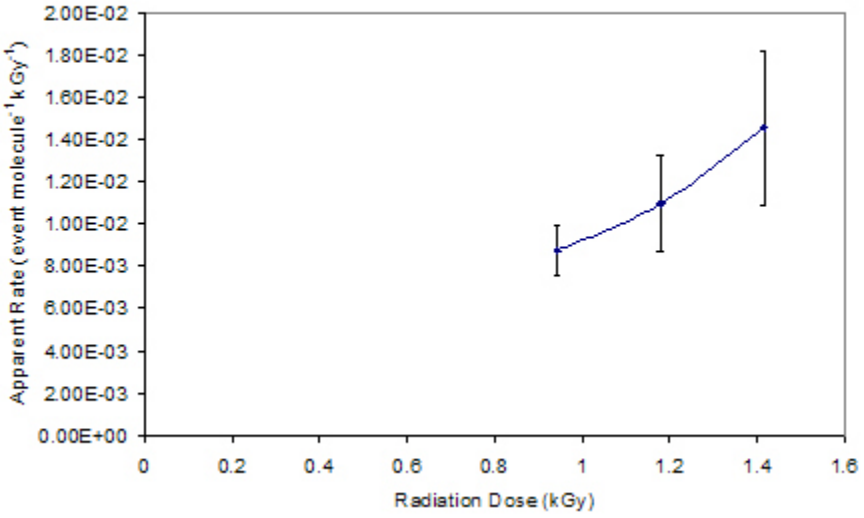


Figure 4.11. The oxidation rate for peptide 105 - 112

Despite the lack of information provided by the kinetics of oxygen addition in this peptide oxidative decarboxylation was detectable. Tandem MS was unable to determine the site(s) of this modification, yet sequence analysis of this peptide suggested that oxidation had to occur on one or more of the four acidic residues of the peptide. The kinetics of oxidative decarboxylation are shown in Figure 4.12. No signal could be detected before 0.708 kGy. Between 0.708 kGy and 1.180 kGy, no increase in the rate of oxidation is detected, indicating that this oxidation target(s) does not experience a conformational change during this time. Between 1.180 kGy and 1.416 kGy, there is a slight increase in the kinetics of oxidation, which suggested that this target(s) does experience a slight increase in accessibility during this phase of irradiation.

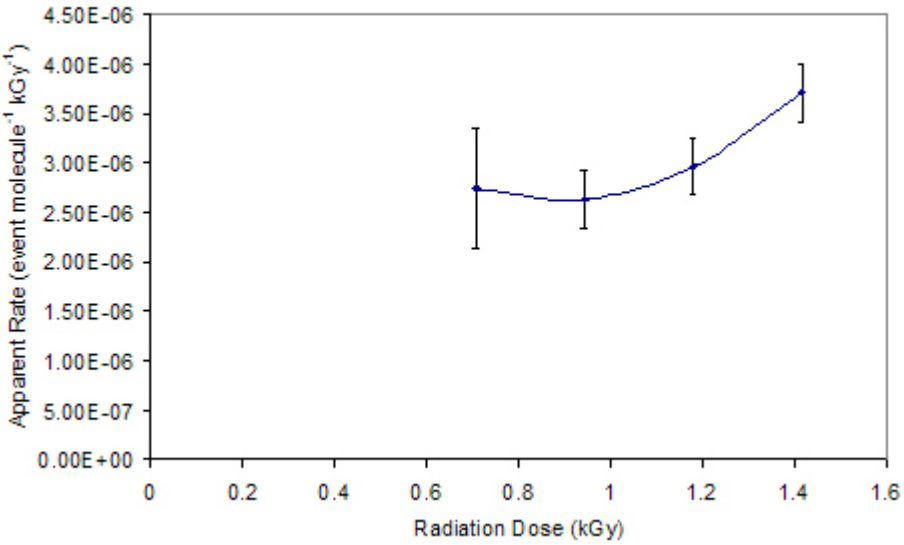


Figure 4.12. The rate of oxidative decarboxylation for peptide 105 - 112

The kinetics of oxidation for peptide 113-117 are shown in Figure 4.13. MS/MS analysis of the oxidized peptide indicated that all detectable oxidation occurred at Asp113. The overall rate of oxidation was very low, in line with the relatively low reactivity of aspartic acid. The kinetics of oxidation indicated a moderate conformational change between 0.236 kGy and 0.472 kGy followed by a region of conformational stability between 0.472 kGy and 0.708 kGy. This peptide did experience a moderate increase in kinetics between 0.708 kGy and 0.944 kGy, indicative of a conformational change. However, this increase was not as drastic as those detected for previous peptides. Examination of the oxidation target for Asp113 (C_β) indicated that this atom is already fairly exposed in the native structure with just under 50% of the solvent accessibility of free aspartic acid (22.07 $Å^2$ versus 46.06 $Å^2$). It is quite possible that at 0.944 kGy Asp113 was fully solvent accessible, and therefore incapable of indicating further conformational change. Thus the lack of increase in the kinetics of oxidation for Asp113 after 0.944 kGy should be interpreted with caution.

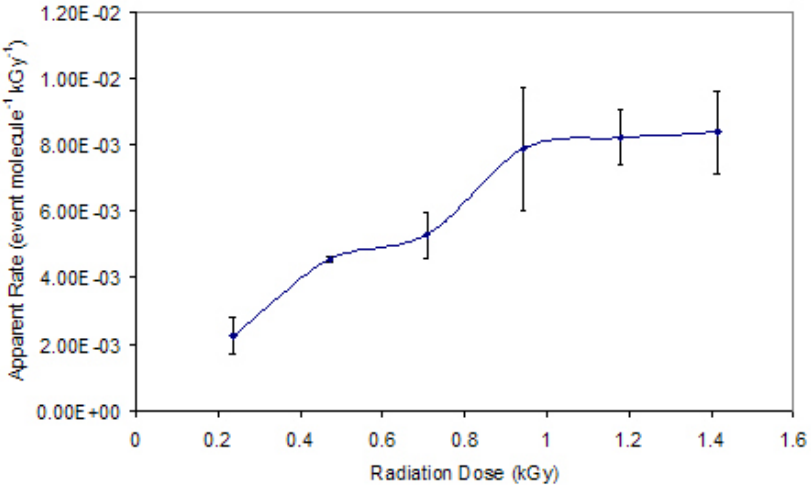


Figure 4.13. The rate of oxidation for peptide 113 - 117

The kinetics of oxidation for peptide 118-122 are shown in Figure 4.14. MS/MS analysis was able to determine that the oxidation was divided between two sites; however, the exact identity of the sites could not be determined due to the low signal of the oxidized peptide. One site was determined to be either Tyr118 or Leu119. The second site was determined to be either Pro120 or Leu121. Analysis of the peptide sequence and structure of peptide 118-122 suggested that the probable targets are Tyr118 and Pro120, with Tyr118 being the more readily oxidized target. The kinetics of this peptide revealed a modest increase in the rate of reaction from 0.236 kGy to 0.472 kGy. From 0.472 kGy to 0.708 kGy no increasewhatsoever was detected, indicating conformational stability over this time. From 0.708 kGy to 0.944 kGy, a modest but significant increase in the kinetics of oxidation was detected, followed by a region of slight decrease from 0.944 kGy to the highest dosage. This

reaction did not proceed beyond the region of linear kinetics; therefore, this slight decrease should be indicative of a small, but real, decrease in the solvent accessibility of the oxidation targets in peptide 118 – 122.

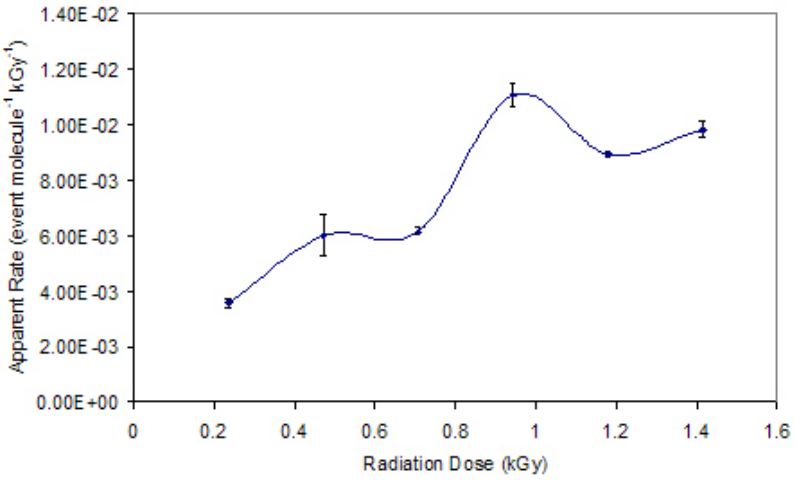


Figure 4.14. The rate of oxidation rate for peptide 118 - 122

Analysis of the oxidation kinetics for each of the peptides presented above revealed that rate of oxidation as a function of radiation dosage fell into three general categories of behavior (Figure 4.15). At lower radiation dosages, less than 0.472 kGy, all peptides exhibited a gradual increase in the rate of oxidation, interpreted as a general destabilization of protein structure that does not have specific structural consequences. This phase of conformational change was termed Stage 1. This was followed by a period of stable rates of oxidation for

almost all peptides measured from 0.472 kGy to 0.708 kGy, indicating a region of conformational stability where additional oxidative damage does not cause marked structural destabilization. From 0.708 kGy to 0.944 kGy, a differential structural response to increased oxidative damage was detected, indicative of a specific conformational change. Certain regions of the protein, illustrated in Figure 4.15 by oxidation targets Met55 and Met60, showed a drastic increase in their rate of oxidation, indicating a dramatic increase in solvent accessibility. Other regions of the protein, illustrated in Figure 4.15 by oxidation target Phe31, did not show a significant increase in the kinetics of oxidation over these radiation dosages, indicating that the conformational change was limited to specific regions of the protein. This phase of conformational change was termed Stage 2. Following this region of conformational change, most peptides showed no significant increase in the kinetics of oxidation between 0.944 kGy and 1.180 kGy, which again indicated a region of relative conformational stability despite additional oxidative damage.

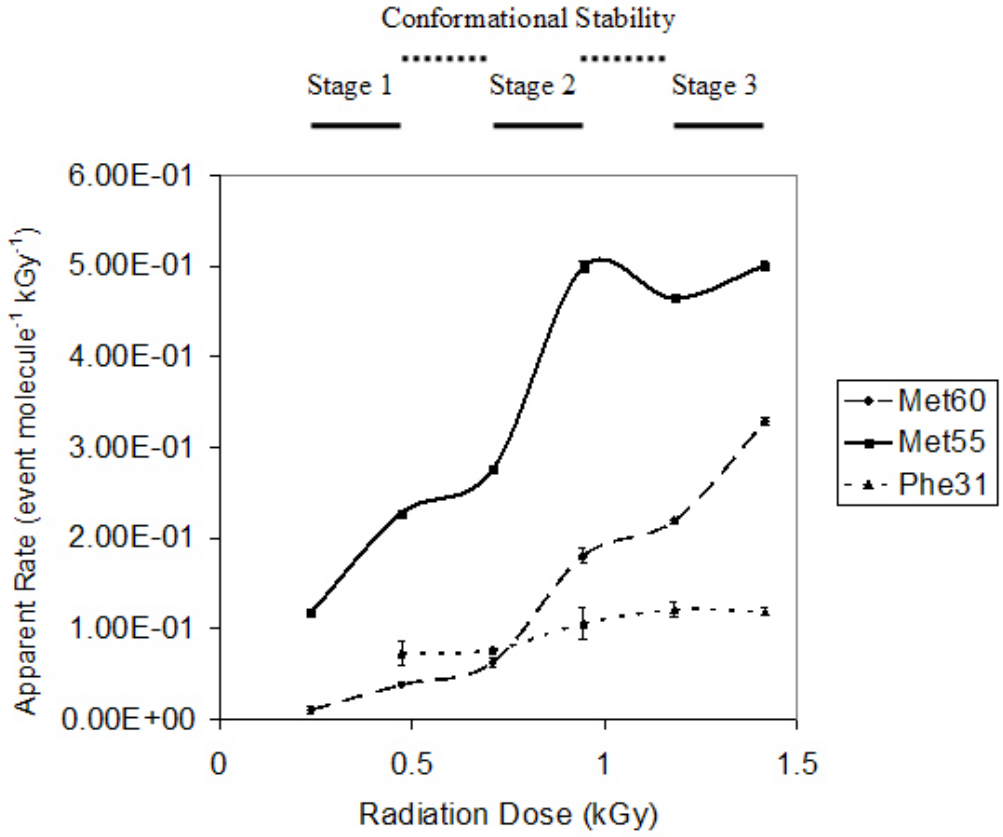


Figure 4.15. Summary of structural response to oxidative damage in Spo0F. The three categories of structural response to oxidative damage, typified by the kinetics of oxidation at three oxidative damage targets. Error bars are shown for all data points and represent one standard deviation.

Finally, between the 1.180 kGy and 1.416 kGy radiation doses, another differential conformational change was detected. Peptides that did not undergo a Stage 2 conformational change (e.g. Phe31) exhibited no further change in the rate of oxidation for these dosages. This indicated that these regions were still resistant to structural changes resulting from oxidative damage. Several peptides that did undergo the Stage 2 conformational change also

showed no increase in the rate of oxidation from 1.180 kGy to 1.416 kGy, illustrated by Met55 in Figure 4.15. However, some of the peptides that experienced a Stage 2 conformational change also showed an increase in the rate of oxidation between these radiation dosages (e.g. Met60), thus indicating that these regions experience an additional increase in solvent accessibility. As stated above, the fact that most regions of Spo0F structure do not experience this additional conformational change (as detected by an increase in the rate of oxidation) indicated that this was a specific conformational change due to oxidative damage, and not a general unfolding of the protein structure.

The sites of protein modification detected and their general response to oxidative damage are listed in Table 4.1. Attempts were made to determine the exact site(s) of modification for each peptide by MS/MS. In cases where the exact sites of modification were not possible to determine by MS/MS (due to low ion signal in the mass spectrum), any partial MS/MS data were considered with the primary sequence and structure of the native conformation of Spo0F. The most likely site of modification was determined based on the known chemical reactivity of the residue, the masses of their likely products, and the average solvent accessibility of the potential targets in the native structure [169-171]. As shown in Table 4.1, solvent accessibility data as a function of radiation dose were measured successfully for eleven peptides in a single experiment, covering many different regions in the structure of Spo0F.

Table 4.1. Increase in the rate of oxidation as a function of radiation dose. Peptides with insufficient signal for oxidized peptides at a particular radiation dose are listed as not determined (ND). *This oxidation target is fully exposed in the native structure, and will not reflect unfolding events. †This oxidative damage is oxidative decarboxylation (loss of 30 Da).

Peptide	Modified Residue	Stage 1 Fold Increase	Post-Stage 1 Fold Increase	Stage 2 Fold Increase	Post-Stage 2 Fold Increase	Stage 3 Fold Increase
6-16	Tyr13*	ND	1.65	0	0	0
17-25	Leu18	4.88	1.30	2.26	0	1.24
26-45	(Tyr28, Phe31)	ND	0	1.38	0	0
46-56	Met55	1.93	1.21	1.81	0.932	1.08
57-67	Met60	3.69	1.63	2.89	1.21	1.50
71-77	Ile72	ND	ND	0	0	0
78-94	Met81, Met89*	ND	0.839	0	0.860	0.881
78-94	Asp88†	ND	3.23	6.44	1.43	1.18
95-104	(Phe102)	3.06	0	2.05	1.44	0
105-112	(Phe106)	ND	ND	ND	0	0
105-112	(Asp107, Asp109, Glu110) †	ND	ND	0	0	1.25
113-117	Asp113	2.03	0	1.49	0	0
118-122	(Tyr118, Pro120)	1.68	0	1.80	0.805	1.10

Discussion

In this work, oxidative surface mapping via mass spectrometry was utilized to independently and simultaneously monitor the solvent accessibility of more than fourteen different oxidation targets in the sporulation inducing protein Spo0F. The kinetics of formation for each of the oxidation products conveys a unique story regarding not only the environment of that particular oxidation target, but also the structural response of Spo0F to oxidative damage. All oxidation targets showed a general increase in the rate of oxidation between 0.236 kGy and 0.472 kGy, indicative of a general destabilization of the protein, followed by a period of conformational stability between 0.472 kGy and 0.708 kGy during which the rate of oxidation does not increase appreciably. However, between 0.708 kGy and 0.944 kGy, the peptides began to differ in their structural response to the process of oxidative damage. Analysis of these data allowed for the elucidation of the specific conformational changes that Spo0F undergoes as a result of oxidative damage, which cannot be determined using a more traditional technique, such as circular dichroism.

Figure 4.16 shows the sites of oxidative damage and the general trend of their kinetics of oxidation from 0.708 kGy to 0.944 kGy. Regions that experience this specific conformational change are grouped near the active site of the protein (exceptions being residues near the C-terminus, which also experienced a modest increase in rate). Interestingly, residues that did not experience an increase in the rate of oxidation during this time are grouped away from the active site, in the structural portion of the protein (except

Met81, as discussed below). These data indicated that the conformational change that occurs between 0.708 kGy and 0.944 kGy exposure involves the loops and the ends of the helices adjacent to the active site, while the structure of the protein distal from the active site (with the exception of the C-terminal tail) remains undisturbed. Additionally, the fact that at least some regions of the active site become even more solvent exposed at higher radiation dosages indicated that the conformational change between 0.708 kGy and 0.944 kGy is not a transition to a wholly disordered state. This conformational change constitutes a specific and controlled unfolding event with a defined intermediate/transition state, rather than a global unravelling of the protein.

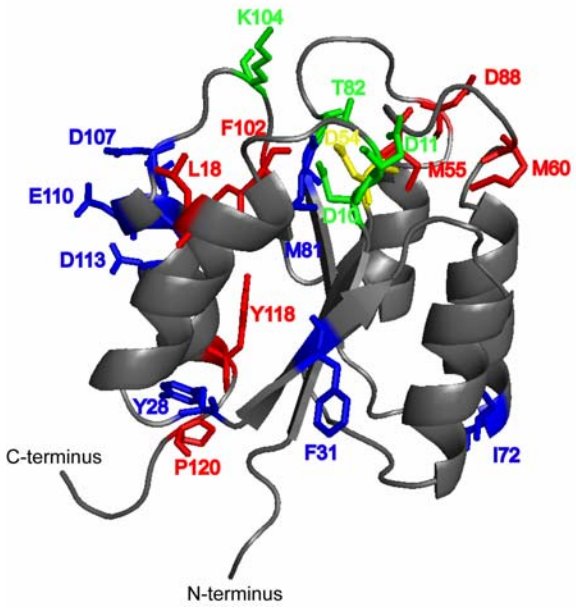


Figure 4.16. Increase in oxidation kinetics from 0.708 kGy to 0.944 kGy plotted on the native structure of Spo0F. Regions showing at least 1.5 – fold increase in oxidation are shown in red, those that did not are shown in blue. Residues essential for phosphorylation (Asp10, Asp11, Thr82, and Lys104) are shown in green. The phosphorylation site, Asp54, is shown in yellow.

In order to lead to such a specific unfolding event, the conformational change experienced from 0.708 kGy to 0.944 kGy exposure must be triggered by oxidative damage at a specific residue. Met60 was oxidized at a very low rate prior to this conformational change; however, after the conformational change it was rapidly oxidized (Figure 4.6). Whatever the triggering oxidation event is, it must have occurred on a sufficient proportion of the protein to account for the increase that occurs from 0.708 kGy to 0.944 kGy exposure. Analysis of the actual amount of oxidation at Met60 after 0.944 kGy compared with the expected amount given no conformational change shows that the trigger site must be at least 12% oxidized at 0.944 kGy exposure to account for the increase in the rate of oxidation of Met60. Examination of the amount of oxidation at each of the other sites shows that only Met55, Met81, and Met89 are modified to a sufficient extent to possibly be the trigger residue.

Another feature of the trigger oxidation event is that it must precede the conformational change and accordingly the kinetics of oxidation at that particular site will not reflect the conformational change itself. Examination of the kinetics of oxidation of Met55 (Figure 4.15) revealed that it does exhibit a marked increase in the rate of oxidation between 0.708 kGy and 0.944 kGy exposure, thus eliminating it from consideration as the likely trigger for conformational change. Because Met81 and Met89 are on the same peptide (residues 78 – 94) any analysis of the kinetics of oxidation will be confused by the fact that the rate reflects the combined rates of oxidation for both oxidation targets. MS/MS analysis of the singly-oxidized and doubly-oxidized peptide conclusively indicates that both residues are oxidized. The intensity of the signal suggests that the oxidation occurred at approximately the same

rate for Met81 and Met89. However, oxidation of sidechains may significantly alter the fragmentation characteristics in both a qualitative and quantitative manner, further confusing attempts to accurately measure the rate of oxidation at Met81 versus Met89. Thus the rate of oxidation of the two sites must be considered together. Therefore, as the oxidation kinetics of this peptide fit the criteria for the Stage 2 trigger site, neither Met81 nor Met89 could be eliminated as the trigger for the conformational change. Examination of the environments of Met81 and Met89 in the structure of Spo0F is quite revealing. While Met89 is almost fully exposed in the native structure, and is not tightly interacting with any other residue, Met81 serves as the core of a tightly packed hydrophobic pocket near the center of the active site. A closer view of this hydrophobic pocket is shown in Figure 4.17. Met81 serves as the center of this pocket, holding together residues from nearby loops and the ends of helices. Previous data also support the conclusion that Met89 is not structurally important, while Met81 is. Alanine scanning mutagenesis indicated that the Met81Ala mutant had decreased enzymatic activity, while the Met89Ala mutant was fully functional. Therefore, we conclude that the conformational change between 0.708 kGy and 0.944 kGy exposure is due to oxidation of Met81 to the sulfoxide.

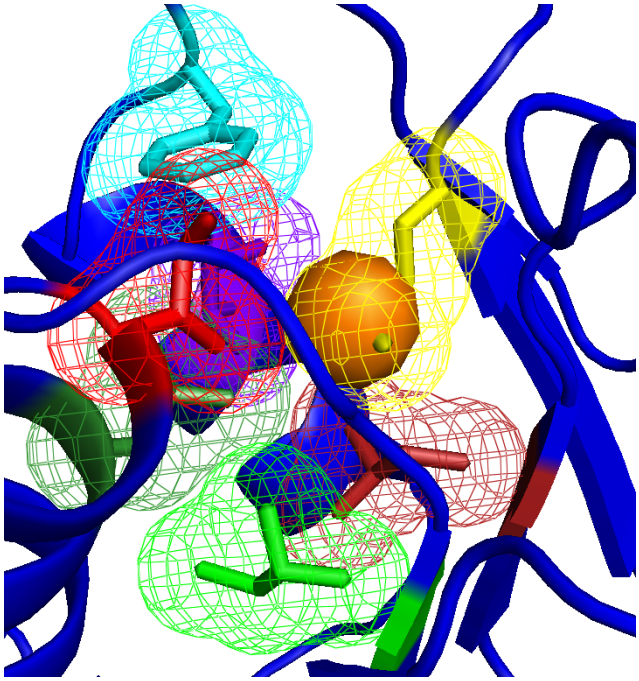


Figure 4.17. Met81 binding pocket. Met81 is shown in yellow, with the sulfur highlighted in orange. Ile8 is shown in green, I15 is colored red, Leu19 is shown in dark green, Leu52 is shown in maroon, Phe106 is shown in cyan, and Ile117 is shown in purple.

Previous NMR data have shown that the β-α loops around the active site are dynamic, which would allow transient access of the hydroxyl radical to Met81. Our data imply a model whereby Met81 is oxidized to the sulfoxide under relatively low radiation dosages (< 0.944 kGy, considerably lower than dosages required to unfold similarly sized globular proteins). The sulfoxide is bulkier than the undamaged methionine, and considerably more polar. Met81=O flips out of the hydrophobic pocket and into the solvent, causing the structural elements that are held together by this hydrophobic pocket to undergo a conformational

change and rearrange to a different pseudo-stable conformation. Further oxidative damage of the protein can disrupt this structure as well (as indicated by increases in the kinetics of oxidation at dosages above 0.944 kGy, such as in Met60 of Figure 4.15), leading to further unfolding in this region, while the rest of the protein remains relatively unchanged.

Spo0F is a member of the CheY superfamily of response regulators. Sequence analysis of the CheY domain shows that, while Met81 is conserved in a significant proportion of family members, it is often substituted for other hydrophobic amino acids. One potential explanation for the conservation of Met in position 81 in some family members but not in others is that a Met in position 81 allows the protein to regulate its function in response to oxidative stress. Perhaps other CheY family members with Met in position 81 act as oxidative stress sensors, modulating cellular response due to specific oxidative damage in the presence of sufficient concentrations of reactive oxygen species.

Future Directions

The above circular dichroism and protein oxidative surface mapping studies have indicated that oxidation of Spo0F results in a two-stage conformation change centered about the active site. Analysis of the kinetics of oxidation at various sites suggested that oxidation of Met81 triggered the observed conformational changes. In order to test this hypothesis several Spo0F Met81 mutants have been created; the first mutant has a leucine substitution at this position while in the second a glutamine has been substituted for methionine. According to

the BLOSUM62 matrix leucine is the most conserved substitution for methionine, and because leucine is relatively insensitive to oxidation when compared to methionine (methionine is approximately five times more reactive with the hydroxyl radical than leucine), it is hypothesized that this mutation will eliminate the specific oxidation-induced conformational changes around the active site while preserving a native structure. In order to determine if the leucine mutant retained native Spo0F structural characteristics, preliminary studies utilizing NMR spectroscopy have been carried out on a ^{15}N labelled protein sample. In Figure 4.18 below, the ^{1}H-^{15}N correlation spectrum of Spo0F Met81Leu is overlaid with the spectrum of wild-type Spo0F. As seen in this figure, the Met81Leu mutant (red) overlays extremely well with wild-type Spo0F (in black), indicating that the leucine mutant retains a nearly wild-type conformation.

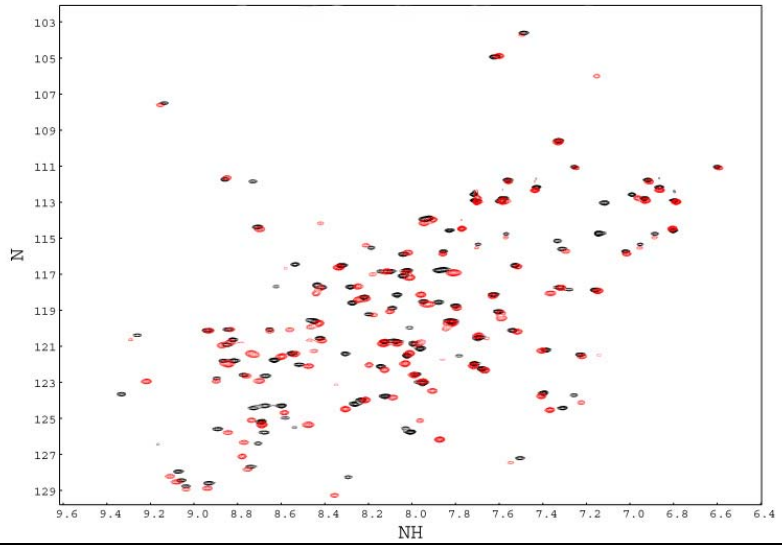


Figure 4.18. ^{1}H-^{15}N HSQC overlay of wild-type Spo0F (black) and Spo0F Met81Leu (red).

While the purpose of the leucine mutant is to determine if oxidation Met81 is necessary to observe the described conformational changes, the purpose of the glutamine mutation is to determine if oxidation at Met81 is sufficient to cause the observed conformational changes. Methionine is a bulky, flexible amino acid, and oxidation of the thioether sulfur to the sulfoxide results in a highly polar S=O bond. Glutamine is similarly bulky, flexible, and is also highly polar. Therefore, a glutamine at this position will provide a provide a structural mimic of oxidation at position 81. This mutant will then be utilized in NMR structural studies to determine the conformation of oxidized Spo0F.

References

1. Pabo, C.O. and R.T. Sauer, *Transcription factors: structural families and principles of DNA recognition.* Annu Rev Biochem, 1992. **61**: p. 1053-95.

2. Huffman, J.L. and R.G. Brennan, *Prokaryotic transcription regulators: more than just the helix-turn-helix motif.* Curr Opin Struct Biol, 2002. **12**(1): p. 98-106.

3. Storz, G. and R. Hengge-Aronis, *Bacterial stress responses*. 2000, Washington, D.C.: ASM Press. xv, 485 p.

4. Trowsdale, J., S.M. Chen, and J.A. Hoch, *Genetic analysis of a class of polymyxin resistant partial revertants of stage O sporulation mutants of Bacillus subtilis: map of the chromosome region near the origin of replication.* Mol Gen Genet, 1979. **173**(1): p. 61-70.

5. Robertson, J.B., et al., *AbrB, a regulator of gene expression in Bacillus, interacts with the transcription initiation regions of a sporulation gene and an antibiotic biosynthesis gene.* Proc Natl Acad Sci U S A, 1989. **86**(21): p. 8457-61.

6. Perego, M., G.B. Spiegelman, and J.A. Hoch, *Structure of the gene for the transition state regulator, abrB: regulator synthesis is controlled by the spo0A sporulation gene in Bacillus subtilis.* Mol Microbiol, 1988. **2**(6): p. 689-99.

7. Strauch, M.A., et al., *The transition state transcription regulator abrB of Bacillus subtilis is a DNA binding protein.* Embo J, 1989. **8**(5): p. 1615-21.

8. Strauch, M.A., et al., *The transition state transcription regulator AbrB of Bacillus subtilis is autoregulated during vegetative growth.* Mol Microbiol, 1989. **3**(9): p. 1203-9.

9. Furbass, R. and M.A. Marahiel, *Mutant analysis of interaction of the Bacillus subtilis transcription regulator AbrB with the antibiotic biosynthesis gene tycA.* FEBS Lett, 1991. **287**(1-2): p. 153-6.

10. Strauch, M.A. and J.A. Hoch, *Transition-state regulators: sentinels of Bacillus subtilis post-exponential gene expression.* Mol Microbiol, 1993. **7**(3): p. 337-42.

11. Fisher, S.H., et al., *Modulation of Bacillus subtilis catabolite repression by transition state regulatory protein AbrB.* J Bacteriol, 1994. **176**(7): p. 1903-12.

12. Strauch, M.A., *In vitro binding affinity of the Bacillus subtilis AbrB protein to six different DNA target regions.* J Bacteriol, 1995. **177**(15): p. 4532-6.

13. Strauch, M.A., *AbrB modulates expression and catabolite repression of a Bacillus subtilis ribose transport operon.* J Bacteriol, 1995. **177**(23): p. 6727-31.
14. Strauch, M.A., *Delineation of AbrB-binding sites on the Bacillus subtilis spo0H, kinB, ftsAZ, and pbpE promoters and use of a derived homology to identify a previously unsuspected binding site in the bsuB1 methylase promote.* J Bacteriol, 1995. **177**(23): p. 6999-7002.

15. Strauch, M.A., *Dissection of the Bacillus subtilis spoOE binding site for the global regulator AbrB reveals smaller recognition elements.* Mol Gen Genet, 1996. **250**(6): p. 742-9.

16. Xu, K., D. Clark, and M.A. Strauch, *Analysis of abrB mutations, mutant proteins, and why abrB does not utilize a perfect consensus in the -35 region of its sigma A promoter.* J Biol Chem, 1996. **271**(5): p. 2621-6.

17. Xu, K. and M.A. Strauch, *In vitro selection of optimal AbrB-binding sites: comparison to known in vivo sites indicates flexibility in AbrB binding and recognition of three-dimensional DNA structures.* Mol Microbiol, 1996. **19**(1): p. 145-58.

18. Vaughn, J.L., et al., *Novel DNA binding domain and genetic regulation model of Bacillus subtilis transition state regulator abrB.* Nat Struct Biol, 2000. **7**(12): p. 1139-46.

19. Vaughn, J.L., et al., *The DNA-binding domain in the Bacillus subtilis transition-state regulator AbrB employs significant motion for promiscuous DNA recognition.* J Mol Biol, 2001. **305**(3): p. 429-39.

20. Phillips, Z.E. and M.A. Strauch, *Role of Cys54 in AbrB multimerization and DNA-binding activity.* FEMS Microbiol Lett, 2001. **203**(2): p. 207-10.

21. Qian, Q., et al., *AbrB is a regulator of the sigma(W) regulon in Bacillus subtilis.* FEMS Microbiol Lett, 2002. **211**(2): p. 219-23.

22. Xu, K. and M.A. Strauch, *DNA-binding activity of amino-terminal domains of the Bacillus subtilis AbrB protein.* J Bacteriol, 2001. **183**(13): p. 4094-8.

23. Saile, E. and T.M. Koehler, *Control of anthrax toxin gene expression by the transition state regulator abrB.* J Bacteriol, 2002. **184**(2): p. 370-80.

24. Shafikhani, S.H., et al., *Postexponential regulation of sin operon expression in Bacillus subtilis.* J Bacteriol, 2002. **184**(2): p. 564-71.

25. Benson, L.M., et al., *Macromolecular assembly of the transition state regulator AbrB in its unbound and complexed states probed by microelectrospray ionization mass spectrometry.* Anal Biochem, 2002. **306**(2): p. 222-7.

26. Cavanagh, J., et al., *Stoichiometries of protein-protein/DNA binding and conformational changes for the transition-state regulator AbrB measured by pseudo cell-size exclusion chromatography-mass spectrometry.* Biochemistry, 2002. **41**(25): p. 7859-65.

27. Bobay, B.G., et al., *Evaluation of the DNA binding tendencies of the transition state regulator AbrB.* Biochemistry, 2004. **43**(51): p. 16106-18.

28. Bobay, B.G., et al., *Revised structure of the AbrB N-terminal domain unifies a diverse superfamily of putative DNA-binding proteins.* FEBS Lett, 2005. **579**(25): p. 5669-74.

29. Hamoen, L.W., et al., *The Bacillus subtilis transition state regulator AbrB binds to the -35 promoter region of comK.* FEMS Microbiol Lett, 2003. **218**(2): p. 299-304.

30. Hamon, M.A., et al., *Identification of AbrB-regulated genes involved in biofilm formation by Bacillus subtilis.* Mol Microbiol, 2004. **52**(3): p. 847-60.

31. Stein, T., *Bacillus subtilis antibiotics: structures, syntheses and specific functions.* Mol Microbiol, 2005. **56**(4): p. 845-57.

32. Coles, M., et al., *AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels.* Structure, 2005. **13**(6): p. 919-28.

33. Strauch, M., et al., *The SpoOA protein of Bacillus subtilis is a repressor of the abrB gene.* Proc Natl Acad Sci U S A, 1990. **87**(5): p. 1801-5.

34. Strauch, M.A. and M. Ayazifar, *Bent DNA is found in some, but not all, regions recognized by the Bacillus subtilis AbrB protein.* Mol Gen Genet, 1995. **246**(6): p. 756-60.

35. Bobay, B.G., et al., *NMR structure of AbhN and comparison with AbrBN: FIRST insights into the DNA binding promiscuity and specificity of AbrB-like transition state regulator proteins.* J Biol Chem, 2006. **281**(30): p. 21399-409.

36. Huang, X. and J.D. Helmann, *Identification of target promoters for the Bacillus subtilis sigma X factor using a consensus-directed search.* J Mol Biol, 1998. **279**(1): p. 165-73.

37. Kunst, F., et al., *The complete genome sequence of the gram-positive bacterium Bacillus subtilis.* Nature, 1997. **390**(6657): p. 249-56.

38. Strauch, M.A., et al., *Abh and AbrB control of Bacillus subtilis antimicrobial gene expression.* J Bacteriol, 2007. **189**(21): p. 7720-32.
39. Bagyan, I., J. Hobot, and S. Cutting, *A compartmentalized regulator of developmental gene expression in Bacillus subtilis.* J Bacteriol, 1996. **178**(15): p. 4500-7.
40. Shcheptov, M., et al., *Characterization of csgA, a new member of the forespore-expressed sigmaG-regulon from Bacillus subtilis.* Gene, 1997. **184**(1): p. 133-40.
41. Igarashi, T. and P. Setlow, *Transcription of the Bacillus subtilis gerK operon, which encodes a spore germinant receptor, and comparison with that of operons encoding other germinant receptors.* J Bacteriol, 2006. **188**(11): p. 4131-6.
42. Wang, S.T., et al., *The forespore line of gene expression in Bacillus subtilis.* J Mol Biol, 2006. **358**(1): p. 16-37.
43. Dong, T.C., S.M. Cutting, and R.J. Lewis, *DNA-binding studies on the Bacillus subtilis transcriptional regulator and AbrB homologue, SpoVT.* FEMS Microbiol Lett, 2004. **233**(2): p. 247-56.
44. Yao, F. and M.A. Strauch, *Independent and interchangeable multimerization domains of the AbrB, Abh, and SpoVT global regulatory proteins.* J Bacteriol, 2005. **187**(18): p. 6354-62.
45. Hoch, J.A. and T.J. Silhavy, *Two-component signal transduction*. 1995, Washington, D.C.: ASM Press. xvi, 488 p.
46. West, A.H. and A.M. Stock, *Histidine kinases and response regulator proteins in two-component signaling systems.* Trends Biochem Sci, 2001. **26**(6): p. 369-76.
47. Stock, A.M., V.L. Robinson, and P.N. Goudreau, *Two-component signal transduction.* Annu Rev Biochem, 2000. **69**: p. 183-215.
48. Parkinson, J.S. and E.C. Kofoid, *Communication modules in bacterial signaling proteins.* Annu Rev Genet, 1992. **26**: p. 71-112.
49. Grebe, T.W. and J.B. Stock, *The histidine protein kinase superfamily.* Adv Microb Physiol, 1999. **41**: p. 139-227.
50. Inouye, M. and R. Dutta, *Histidine kinases in signal transduction*. 2003, San Diego, Calif.: Academic Press. xviii, 520 p.
51. Hoch, J.A., *Two-component and phosphorelay signal transduction.* Curr Opin Microbiol, 2000. **3**(2): p. 165-70.

52. Khorchid, A. and M. Ikura, *Bacterial histidine kinase as signal sensor and transducer.* Int J Biochem Cell Biol, 2006. **38**(3): p. 307-12.
53. Egger, L.A., H. Park, and M. Inouye, *Signal transduction via the histidyl-aspartyl phosphorelay.* Genes Cells, 1997. **2**(3): p. 167-84.

54. Appleby, J.L. and R.B. Bourret, *Proposed signal transduction role for conserved CheY residue Thr87, a member of the response regulator active-site quintet.* J Bacteriol, 1998. **180**(14): p. 3563-9.

55. Lukat, G.S., et al., *Roles of the highly conserved aspartate and lysine residues in the response regulator of bacterial chemotaxis.* J Biol Chem, 1991. **266**(13): p. 8348-54.

56. Zhu, X., et al., *Crystal structures of CheY mutants Y106W and T87I/Y106W. CheY activation correlates with movement of residue 106.* J Biol Chem, 1997. **272**(8): p. 5000-6.

57. Hoch, J.A. and K.I. Varughese, *Keeping signals straight in phosphorelay signal transduction.* J Bacteriol, 2001. **183**(17): p. 4941-9.

58. Galperin, M.Y., *Structural classification of bacterial response regulators: diversity of output domains and domain combinations.* J Bacteriol, 2006. **188**(12): p. 4169-82.

59. Hoch, J.A., *Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis.* Annu Rev Microbiol, 1993. **47**: p. 441-65.

60. Trach, K., et al., *Control of the initiation of sporulation in Bacillus subtilis by a phosphorelay.* Res Microbiol, 1991. **142**(7-8): p. 815-23.

61. Burbulys, D., K.A. Trach, and J.A. Hoch, *Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay.* Cell, 1991. **64**(3): p. 545-52.

62. Hilbert, D.W. and P.J. Piggot, *Compartmentalization of gene expression during Bacillus subtilis spore formation.* Microbiol Mol Biol Rev, 2004. **68**(2): p. 234-62.

63. Li, Z., F. Di Donato, and P.J. Piggot, *Compartmentalization of gene expression during sporulation of Bacillus subtilis is compromised in mutants blocked at stage III of sporulation.* J Bacteriol, 2004. **186**(7): p. 2221-3.

64. Stragier, P. and R. Losick, *Molecular genetics of sporulation in Bacillus subtilis.* Annu Rev Genet, 1996. **30**: p. 297-41.

65. Phillips, Z.E. and M.A. Strauch, *Bacillus subtilis sporulation and stationary phase gene expression.* Cell Mol Life Sci, 2002. **59**(3): p. 392-402.

66. Errington, J., *Regulation of endospore formation in Bacillus subtilis.* Nat Rev Microbiol, 2003. **1**(2): p. 117-26.

67. Piggot, P.J. and D.W. Hilbert, *Sporulation of Bacillus subtilis.* Curr Opin Microbiol, 2004. **7**(6): p. 579-86.

68. Yudkin, M.D. and J. Clarkson, *Differential gene expression in genetically identical sister cells: the initiation of sporulation in Bacillus subtilis.* Mol Microbiol, 2005. **56**(3): p. 578-89.

69. Grossman, A.D., *Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis.* Annu Rev Genet, 1995. **29**: p. 477-508.

70. Jiang, M., et al., *Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis.* Mol Microbiol, 2000. **38**(3): p. 535-42.

71. Hahn, J., M. Roggiani, and D. Dubnau, *The major role of Spo0A in genetic competence is to downregulate abrB, an essential competence gene.* J Bacteriol, 1995. **177**(12): p. 3601-5.

72. Fujita, M., J.E. Gonzalez-Pastor, and R. Losick, *High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis.* J Bacteriol, 2005. **187**(4): p. 1357-68.

73. Fawcett, P., et al., *The transcriptional profile of early to middle sporulation in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2000. **97**(14): p. 8063-8.

74. Perego, M., P. Glaser, and J.A. Hoch, *Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in Bacillus subtilis.* Mol Microbiol, 1996. **19**(6): p. 1151-7.

75. Perego, M., *A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of Bacillus subtilis.* Mol Microbiol, 2001. **42**(1): p. 133-43.

76. Perego, M., et al., *Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in B. subtilis.* Cell, 1994. **79**(6): p. 1047-55.

77. Driks, A., *Overview: Development in bacteria: spore formation in Bacillus subtilis.* Cell Mol Life Sci, 2002. **59**(3): p. 389-91.

78. Alsaker, K.V., T.R. Spitzer, and E.T. Papoutsakis, *Transcriptional analysis of spo0A overexpression in Clostridium acetobutylicum and its effect on the cell's response to butanol stress.* J Bacteriol, 2004. **186**(7): p. 1959-71.
79. Grandvalet, C., M. Gominet, and D. Lereclus, *Identification of genes involved in the activation of the Bacillus thuringiensis inhA metalloprotease gene at the onset of sporulation.* Microbiology, 2001. **147**(Pt 7): p. 1805-13.

80. Strauch, M.A., et al., *The DNA-binding specificity of the Bacillus anthracis AbrB protein.* Microbiology, 2005. **151**(Pt 6): p. 1751-9.

81. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information.* J Am Chem Soc, 2003. **125**(7): p. 1731-7.

82. Grzesiek, S. and A. Bax, *Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins.* J Biomol NMR, 1993. **3**(2): p. 185-204.

83. Ikura, M., L.E. Kay, and A. Bax, *A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin.* Biochemistry, 1990. **29**(19): p. 4659-67.

84. Logan, T.M., et al., *Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments.* FEBS Lett, 1992. **314**(3): p. 413-8.

85. Logan, T.M., et al., *A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments.* J Biomol NMR, 1993. **3**(2): p. 225-31.

86. Montelione, G.T., S.D. Emerson, and B.A. Lyons, *A general approach for determining scalar coupling constants in polypeptides and proteins.* Biopolymers, 1992. **32**(4): p. 327-34.

87. Cavanagh, J., Faculty Publication Collection (North Carolina State University), and Wartime Classes Golden Anniversary Endowment, *Protein NMR spectroscopy : principles and practice*. 2nd ed. 2007, Amsterdam ; Boston: Academic Press. xxv, 885 p.

88. Cornilescu, G., F. Delaglio, and A. Bax, *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* J Biomol NMR, 1999. **13**(3): p. 289-302.

89. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

90. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol Biol, 2004. **278**: p. 313-52.

91. Chou, J.J., et al., *A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles.* J Biomol NMR, 2001. **21**(4): p. 377-82.

92. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.

93. Linge, J.P., S.I. O'Donoghue, and M. Nilges, *Automated assignment of ambiguous nuclear overhauser effects with ARIA.* Methods Enzymol, 2001. **339**: p. 71-90.

94. Farrow, N.A., et al., *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry, 1994. **33**(19): p. 5984-6003.

95. Skelton, N.J., Palmer III, A.G., Akke, M., Kordel, J., Rance, M., and Chazin, W.J., *Practical Aspects of Two-Dimensional Proton-Detected 15N spin Relaxation Measurements.* Journal of Magnetic Resonance, 1993. **102**(ser. B.): p. 253-264.

96. Theobald, D.L. and D.S. Wuttke, *THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures.* Bioinformatics, 2006. **22**(17): p. 2171-2.

97. Jarymowycz, V.A. and M.J. Stone, *Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences.* Chem Rev, 2006. **106**(5): p. 1624-71.

98. d'Auvergne, E.J. and P.R. Gooley, *Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces.* J Biomol NMR, 2008. **40**(2): p. 107-19.

99. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M., *Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins.* Journal of the American Chemical Society, 1990. **112**: p. 4989-4991.

100. Lipari, G.a.S., A., *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1 Theory and range of validity.* Journal of the American Chemistry Society, 1982. **104**(4546-4559).

101. Lipari, G.a.S., A., *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.* Journal of the American Chemistry Society, 1982. **104**: p. 4559-4570.
102. Clore, G.M., et al., *Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy.* Biochemistry, 1990. **29**(32): p. 7387-401.

103. Mandel, A.M., M. Akke, and A.G. Palmer, 3rd, *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme.* J Mol Biol, 1995. **246**(1): p. 144-63.

104. d'Auvergne, E.J. and P.R. Gooley, *Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data.* J Biomol NMR, 2006. **35**(2): p. 117-35.

105. d'Auvergne, E.J. and P.R. Gooley, *The use of model selection in the model-free analysis of protein dynamics.* J Biomol NMR, 2003. **25**(1): p. 25-39.

106. Sali, A., et al., *Evaluation of comparative protein modeling by MODELLER.* Proteins, 1995. **23**(3): p. 318-26.

107. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., *PROCHECK: a program to check the stereochemical quality of protein structures.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

108. Hubbard, S.J., Thorton, J.M., *NACCESS*. 1993, Department of Biochemistry and Molecular Biology, University College London.

109. van Dijk, M., et al., *Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility.* Nucleic Acids Res, 2006. **34**(11): p. 3317-25.

110. Lu, X.J. and W.K. Olson, *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures.* Nucleic Acids Res, 2003. **31**(17): p. 5108-21.

111. Kneller, J.M., M. Lu, and C. Bracken, *An effective method for the discrimination of motional anisotropy and chemical exchange.* J Am Chem Soc, 2002. **124**(9): p. 1852-3.

112. Dickerson, R.E. and T.K. Chiu, *Helix bending as a factor in protein/DNA recognition.* Biopolymers, 1997. **44**(4): p. 361-403.

113. Olson, W.K., et al., *DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11163-8.

114. Rhodes, D., et al., *Towards an understanding of protein-DNA recognition.* Philos Trans R Soc Lond B Biol Sci, 1996. **351**(1339): p. 501-9.
115. Suzuki, M., N. Yagi, and M. Gerstein, *DNA recognition and superstructure formation by helix-turn-helix proteins.* Protein Eng, 1995. **8**(4): p. 329-38.

116. Travers, A.A., *The structural basis of DNA flexibility.* Philos Transact A Math Phys Eng Sci, 2004. **362**(1820): p. 1423-38.

117. Coles, M., et al., *Common evolutionary origin of swapped-hairpin and double-psi beta barrels.* Structure, 2006. **14**(10): p. 1489-98.

118. Berman, H.M., *Crystal studies of B-DNA: the answers and the questions.* Biopolymers, 1997. **44**(1): p. 23-44.

119. Schneider, B., Neidle, S., Berman H.M., *Conformations of the sugar-phosphate backbone in helical DNA crystal structures.* Biopolymers, 1997. **42**(1): p. 113-124.

120. Ouzounis, C.A., et al., *Classification schemes for protein structure and function.* Nat Rev Genet, 2003. **4**(7): p. 508-19.

121. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-74.

122. Meeks, J.C., et al., *An overview of the genome of Nostoc punctiforme, a multicellular, symbiotic cyanobacterium.* Photosynth Res, 2001. **70**(1): p. 85-106.

123. Sola, M., et al., *Three-dimensional crystal structure of the transcription factor PhoB receiver domain.* J Mol Biol, 1999. **285**(2): p. 675-87.

124. Gouet, P., et al., *Structural transitions in the FixJ receiver domain.* Structure, 1999. **7**(12): p. 1517-26.

125. Muller-Dieckmann, H.J., A.A. Grantz, and S.H. Kim, *The structure of the signal receiver domain of the Arabidopsis thaliana ethylene receptor ETR1.* Structure, 1999. **7**(12): p. 1547-56.

126. Im, Y.J., et al., *Crystal structure of a cyanobacterial phytochrome response regulator.* Protein Sci, 2002. **11**(3): p. 614-24.

127. Simonovic, M. and K. Volz, *A distinct meta-active conformation in the 1.1-A resolution structure of wild-type ApoCheY.* J Biol Chem, 2001. **276**(31): p. 28637-40.

128. Benda, C., et al., *Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators.* Biophys J, 2004. **87**(1): p. 476-87.
129. Buckler, D.R., Y. Zhou, and A.M. Stock, *Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima.* Structure, 2002. **10**(2): p. 153-64.

130. Guillet, V., et al., *Crystallographic and biochemical studies of DivK reveal novel features of an essential response regulator in Caulobacter crescentus.* J Biol Chem, 2002. **277**(44): p. 42003-10.

131. Birck, C., et al., *The crystal structure of the phosphorylation domain in PhoP reveals a functional tandem association mediated by an asymmetric interface.* J Bacteriol, 2003. **185**(1): p. 254-61.

132. Lee, S.Y., et al., *Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains.* Genes Dev, 2003. **17**(20): p. 2552-63.

133. Usher, K.C., et al., *Crystal structures of CheY from Thermotoga maritima do not support conventional explanations for the structural basis of enhanced thermostability.* Protein Sci, 1998. **7**(2): p. 403-12.

134. Chan, C., et al., *Structural basis of activity and allosteric control of diguanylate cyclase.* Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17084-9.

135. Toro-Roman, A., T.R. Mack, and A.M. Stock, *Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face.* J Mol Biol, 2005. **349**(1): p. 11-26.

136. Feher, V.A., et al., *High-resolution NMR structure and backbone dynamics of the Bacillus subtilis response regulator, Spo0F: implications for phosphorylation and molecular recognition.* Biochemistry, 1997. **36**(33): p. 10015-25.

137. Fabret, C., V.A. Feher, and J.A. Hoch, *Two-component signal transduction in Bacillus subtilis: how one organism sees its world.* J Bacteriol, 1999. **181**(7): p. 1975-83.

138. Zapf, J., et al., *A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction.* Structure, 2000. **8**(8): p. 851-62.

139. Tzeng, Y.L. and J.A. Hoch, *Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis.* J Mol Biol, 1997. **272**(2): p. 200-12.

140. Xu, Q., S.W. Porter, and A.H. West, *The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems.* Structure, 2003. **11**(12): p. 1569-81.

141. Zhao, R., et al., *Structure and catalytic mechanism of the E. coli chemotaxis phosphatase CheZ.* Nat Struct Biol, 2002. **9**(8): p. 570-5.

142. Ohta, N. and A. Newton, *The core dimerization domains of histidine kinases contain recognition specificity for the cognate response regulator.* J Bacteriol, 2003. **185**(15): p. 4424-31.

143. Kojetin, D.J., R.J. Thompson, and J. Cavanagh, *Sub-classification of response regulators using the surface characteristics of their receiver domains.* FEBS Lett, 2003. **554**(3): p. 231-6.

144. Chen, C.Y., et al., *Comparative genome analysis of Vibrio vulnificus, a marine pathogen.* Genome Res, 2003. **13**(12): p. 2577-87.

145. Hubbard, T.J., et al., *SCOP: a structural classification of proteins database.* Nucleic Acids Res, 1997. **25**(1): p. 236-9.

146. Gough, J., et al., *Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure.* J Mol Biol, 2001. **313**(4): p. 903-19.

147. Benson, D.A., et al., *GenBank.* Nucleic Acids Res, 2005. **33**(Database issue): p. D34-8.

148. Chapman, B., and Chang, J., *Biopython: Python tools for computational biology.* ACM SIGBIO Newslett., 2000. **20**: p. 15–19.

149. Wang, G. and R.L. Dunbrack, Jr., *PISCES: a protein sequence culling server.* Bioinformatics, 2003. **19**(12): p. 1589-91.

150. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.

151. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions.* Protein Sci, 1993. **2**(9): p. 1511-9.

152. Eisenberg, D., R. Luthy, and J.U. Bowie, *VERIFY3D: assessment of protein models with three-dimensional profiles.* Methods Enzymol, 1997. **277**: p. 396-404.

153. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.

154. DeLano, W.L., *ThePyMOLmolecular graphics system.* 2002, DeLano Scientific: PaloAlto,CA.

155. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.

156. Wolfenden, R., et al., *Affinities of amino acid side chains for solvent water.* Biochemistry, 1981. **20**(4): p. 849-55.

157. Sahoo, S., et al., *Intracellular reactive oxygen species mediate suppression of sporulation in Bacillus subtilis under shear stress.* Biotechnol Bioeng, 2004. **87**(1): p. 81-9.

158. Tzeng, Y.L., et al., *Characterization of interactions between a two-component response regulator, Spo0F, and its phosphatase, RapB.* Biochemistry, 1998. **37**(47): p. 16538-45.

159. Feher, V.A., et al., *1H, 15N, and 13C backbone chemical shift assignments, secondary structure, and magnesium-binding characteristics of the Bacillus subtilis response regulator, Spo0F, determined by heteronuclear high-resolution NMR.* Protein Sci, 1995. **4**(9): p. 1801-14.

160. Feher, V.A. and J. Cavanagh, *Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F.* Nature, 1999. **400**(6741): p. 289-93.

161. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochem Biophys Res Commun, 2001. **287**(3): p. 614-21.

162. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry.* Anal Chem, 2004. **76**(3): p. 672-83.

163. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry.* Anal Biochem, 2003. **313**(2): p. 216-25.

164. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* J Am Soc Mass Spectrom, 2005. **16**(12): p. 2057-63.
165. Goldsmith, S.C., et al., *Synchrotron protein footprinting: a technique to investigate protein-protein interactions.* J Biomol Struct Dyn, 2001. **19**(3): p. 405-18.

166. Maleknia, S.D., et al., *Determination of macromolecular folding and structure by synchrotron x-ray radiolysis techniques.* Anal Biochem, 2001. **289**(2): p. 103-15.

167. Andrade, M.A., et al., *Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network.* Protein Eng, 1993. **6**(4): p. 383-90.

168. Sharp, J.S. and K.B. Tomer, *Analysis of the oxidative damage-induced conformational changes of apo- and holocalmodulin by dose-dependent protein oxidative surface mapping.* Biophys J, 2007. **92**(5): p. 1682-92.

169. Buxton, G.V.G., C. L.; Helman, W. P.; Ross, A. B., *Critical-Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen-Atoms and Hydroxyl Radicals (.Oh/.O-) in Aqueous-Solution.* Journal of Physical and Chemical Reference 1988. **17**(2): p. 513-886.

170. Garrison, W.M., *Reaction-Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins.* Chemical Reviews, 1988. **87**(2): p. 381-398.

171. Xu, G. and M.R. Chance, *Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting.* Anal Chem, 2005. **77**(14): p. 4549-55.

172. Guan, J.Q. and M.R. Chance, *Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry.* Trends Biochem Sci, 2005. **30**(10): p. 583-92.

Appendices

Appendix A

NMR Spectroscopy

In order to fully understand the function of a biological macromolecule, it is critical to first determine the high-resolution three dimensional structure. There are currently two techniques to obtain such detailed structural information at the atomic level, X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. The advantage of NMR spectroscopy over X-ray crystallography lies in the fact that an NMR structure is a solution structure, and NMR spectroscopy allows for the determination of dynamic information on a variety of biologically relevant timescales. The following section provides a brief overview of the theory and techniques of NMR spectroscopy as it relates to the work described in this dissertation.

General NMR Theory

The phenomenon that gives rise to NMR spectroscopy is the property of *nuclear spin* [1]. The spin number, *I*, depends on the number of protons and neutrons in the nucleus of an atom. There are several rules for determining the net spin of a nucleus: i) for even mass nuclei with an even number of neutrons spin is zero, ii) for even mass nuclei with an odd number of neutrons spin is an integer value, iii) for nuclei with an odd mass spin is ½ integer value. Because the phenomenon that gives rise to NMR relies upon the existence of nuclear

spin, those nuclei in category (i) above are NMR inactive. For NMR spectroscopy of proteins, the most useful nuclei are ^{1}H, ^{15}N, and ^{13}C, which following rule (iii) above are spin ½. In the presence of an external magnetic field (B_0), these nuclei will orient the axis of their dipole moment in certain orientations, corresponding to discrete energy levels given by the relationship $2I + 1$. Thus spin ½ nuclei can occupy two energy levels corresponding to the quantum numbers m = +½ and -½ . In the presence of the external magnetic field, the magnetic dipole oriented with the field (m = +½) has the lowest energy, α, while the magnetic dipole oriented against the field (m = -½) has the highest energy, β (Figure A.1). In the absence of an external magnetic field, the energy levels are of equal energy, or degenerate.

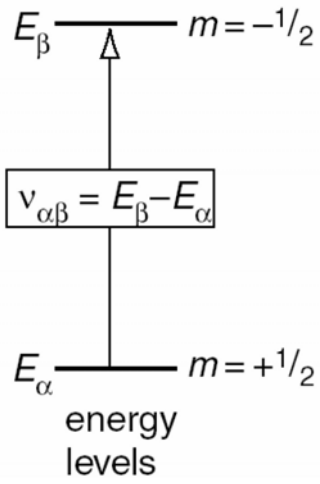


Figure A.1. Energy transition between α and β states. For spins of ±½ , the frequency needed to complete a transition from the α to β state is given by $\nu\alpha_\beta$

In the presence of an external magnetic field, a transition between the lower energy and the higher energy level is possible by the application of a radiofrequency pulse of a specific

wavelength, $\nu_{\alpha\beta}$. This wavelength corresponds to the energy difference between the α and β states, termed the transition energy, ΔE. The energy of any particular energy level is given by:

$$E = -\gamma m \hbar B_0 \quad \text{(Equation A1)}$$

where γ is the gyromagnetic ratio, m is the quantum number corresponding to the energy level, $\hbar$ is Planck's constant divided by 2π, and B_0 is the external magnetic field (in Tesla). Following this, the transition energy between the α and β states in a spin ½ nucleus is given by:

$$\Delta E = \gamma \hbar B_0 \quad \text{(Equation A2)}$$

As can be seen from the above equation, the transition energy between the α and β states is proportional to the both the gyromagnetic ratio and the strength of the magnetic field B_0. In the presence of an external magnetic field, an NMR sensitive nuclei will precess about the z-axis defined by the external magnetic field at a characteristic frequency, termed the Larmor frequency [2]. The Larmor frequency can be determined by the following equation:

$$\omega_0 = -\gamma B_0 \quad \text{(Equation A3)}$$

To this point we have considered the case of a single spin. In biomacromolecules such as proteins, there exist a great number of such nuclei. The distribution of these nuclei between

the low and high energy levels at equilibrium in a magnetic field is given by the Boltzmann distribution. The equation for the Boltzmann distribution is below, followed by a graphical representation in Figure A.2:

$$N_\alpha/N_\beta = e^{-\Delta E/\kappa T} \quad \text{(Equation A4)}$$

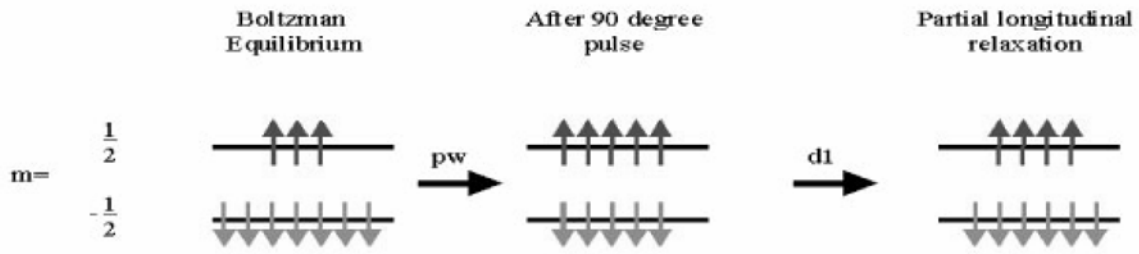


Figure A.2. Boltzmann distribution of nuclei in the excited (β) and lower energy (α) states at equilibrium, after a 90° radiofrequency pulse, and after relaxation

where N_α is the population of nuclei in the high energy state, N_β is that of nuclei in the high energy state, κ is the Boltzmann constant, T is the temperature in Kelvin and ΔE is transition energy between the α and β states. At equilibrium there is a slight excess of nuclear spins in the low energy state, which gives rise to a bulk magnetic moment, M, parallel to the external magic field along the z-axis. This magnetic moment precesses about the main field at the Larmor frequency described above. A graphical representation is shown below (Figure A.3).

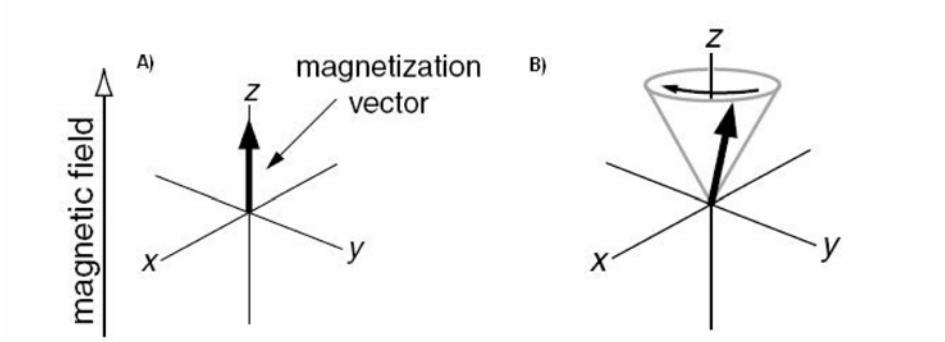


Figure A.3. Bulk magnetization. *A*, Nuclei have a net magnetization along the direction of the magnetic field (z-axis) at equilibrium. *B*, A radiofrequency pulse can tilt the magnetization vector such that it creates an angle to the z-axis, causing the vector to rotate about the z-axis at a rate determined by the Larmor frequency

The application of a radiofrequency pulse (B1) in the x,y plane can flip the bulk magnetization from the z-axis to the x,y-axis. The appropriate pulse width (length of time) can lead to a 90^o pulse, which results in an equalization of the α and β states. This results in identical superpositions of energy phases and amplitudes of the states. This superposition is a type of communication between the energy levels, which is termed coherence. Multiple pulse sequences at specific radio frequencies can transfer magnetization from one atom to another, i.e. H^N to N, through coherence. The magnetization in the x,y plane is termed transverse magnetization and is subject to the static magnetic field of the spectrometer (B_0). This force leads to precession of the spin at the Larmor frequency of the specific atom. If unperturbed, the spin will eventually relax back to equilibrium. From this relaxation back to equilibrium, two rate constants can be obtained: i) $1/T_1$, called the longitudinal or spin-lattice relaxation rate, is the rate at which magnetization builds up parallel to the B_0; ii) $1/T_2$, called

transverse of spin-spin relaxation rate, is the rate at which transverse magnetization in the x,y plane is lost. As a consequence of the fact that the processes that contribute to T_1 also contribute to T_2, by definition T_2 can never be longer than T_1.

The precession of the bulk magnetization about the z-axis of the external magnetic field is what is detected in the NMR experiment. The return to equilibrium, from the x,y plane to z-axis generates an electrical current in a coil aligned with the x-axis of the sample. This recorded signal is referred to as the free-induction decay (FID). The FID consists of a complex mixture of decaying cosine waves, which represents the sum of the resonance frequencies of the NMR observable nuclei in the sample. The FID is then converted from the time domain to the frequency domain by use of the Fourier transformation. This is depicted graphically below (Figure A.4).

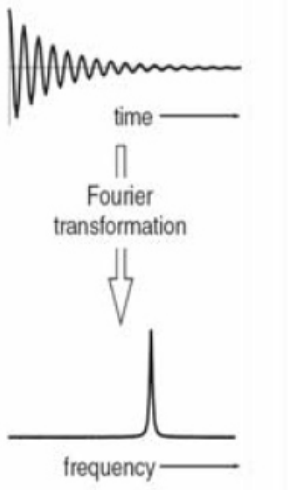


Figure A.4. The FID. Schematic representation of the application of a RF pulse to a system at equilibrium, the collection of a FID, and the Fourier transformation of the FID from the time domain to the frequency domain, resulting in a single peak in the NMR spectrum.

The types of information can be obtained by NMR spectroscopy that relate to the protein being studied. These include chemical shifts, scalar couplings, and dipolar couplings. When a sample is exposed to an external magnetic field, the electrons of all atoms in the sample circulate about the direction of the applied field. This circulation causes a slight secondary magnetic field. Thus, the net magnetic field at a particular nucleus depends on the static magnetic field and the local secondary fields. The effect of these secondary magnetic fields is called nuclear shielding and can either enhance or diminish the effect of the static magnetic field. The chemical shift, δ, is derived from the comparison of the frequency of a sample nucleus to the frequency of a reference compound:

$$\delta = (\nu_r - \nu_s)/ \nu_o \quad \text{(Equation A5)}$$

where ν_s is the frequency of the sample nuclei, ν_r is the frequency of the reference compound, and ν_o is the frequency of excitation that corresponds to the magnetic field. The chemical shift has the units of parts per million (ppm). ν_s is unique to each nuclei due effects arising from the nuclear shielding effect of the local electronic environment, which as stated above can either enhance or diminish the main field at each nuclei. Thus, it is through the unique chemical shifts that we are able to define unique three-dimensional positions for all of the various nuclei in a protein.

Scalar couplings, also termed J-couplings or through-bond interactions, arise between two

spin ½ nuclei and are mediated by the electrons involved in covalent bonds linking the nuclei, thus allowing for the transfer of magnetization between the two spin ½ nuclei. Scalar coupling gives rise to multiplets in the spectrum where, for example, if two spin ½ nuclei are scalar coupled, the resonance from each spin will split symmetrically about the chemical shift, forming a doublet. In this simplistic case, each doublet is split by the same amount, referred to as the coupling constant. The amount of splitting is independent of the magnetic field and is given in units of Hz. The further the distance between the atoms, the smaller the value of the coupling. There is an upper limit of four bonds between atoms, beyond which splitting is too small to discern. Examples of these couplings along the polypeptide chain are shown below.

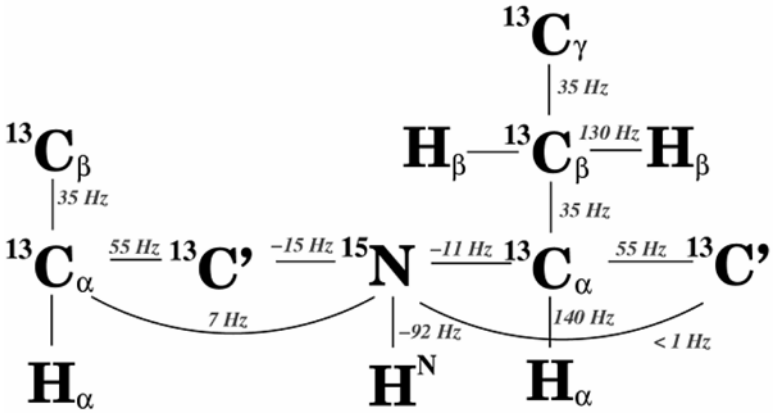


Figure A.5. One bond heteronuclear scalar couplings along the polypeptide chain which are utilized in 3D and 4D NMR experiments.

Dipolar couplings, or through-space interactions, are the third type of information that can be determined through the use of NMR. Dipolar coupling occurs between nuclei that are close

in three-dimensional space and is dependent on the distance between the nuclei, as governed by r^{-6}, where r is the distance between the nuclei. This observed phenomenon is called the nuclear Overhauser effect (NOE), which is the foundation for a whole suite of NOESY experiments, which provide the central foundation with which NMR spectroscopy can be used to develop high-resolution three-dimensional structures of molecules. Dependence on the inverse six power of the distance between the nuclei causes NOE signal intensity to diminish rapidly with increasing distance, leading to an upper limit of detection of approximately 5 Å.

Protein Labeling Strategies

In order to study proteins using NMR spectroscopy we must first isotopically label the protein with NMR active nuclei. Hydrogen is high in natural abundance and is incorporated by the cells naturally, so the growth media need not be enriched with protons. However, the most common isotopes of nitrogen and carbon (^{14}N and ^{12}C) are not NMR active nuclei, and as such, the growth media must be enriched with the NMR active isotopes of these elements in order to achieve uniform labeling of proteins. After proper conditions for growth and purification are determined using unlabeled protein expressed by cells grown in a rich media such as LB, cells can be grown in minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as the sole sources of nitrogen and carbon. Once proteins are uniformly labeled, they may be studied using NMR experiments described below.

Protein Backbone and Sidechain Resonance Assignment

In order to structure a protein using NMR, the backbone and sidechain resonances must first be assigned. This will allow for the proper analysis of NOE experiments which will provide the geometric restraints that will fold the linear amino acid sequence of a protein into the final three-dimensional tertiary fold. Unfortunately, the NMR frequency distribution of nuclei present in a biomacromolecules is small, and in a typical protein there are several hundreds of NMR active nuclei that give rise to detectable resonances. Because of this, in any sizeable protein there will be a great deal of resonance overlap. To overcome this, multidimensional heteronuclear NMR techniques were developed, allowing for the unambiguous assignment of resonances to particular nuclei in the protein.

This process is begun by running a ^{15}N-HSQC (heteronuclear spin quantum correlation) experiment, which produces a resonance peak between every amide nitrogen (dimension 1) and amide proton (dimension 2) to which is attached. The primary purpose of this experiment is to determine if the appropriate number of cross peaks are visualized for the expected number of N-H groups in the protein of interest. Once this has been determined, the ^{15}N – ^{1}H information provides the framework which we build upon in a series of 3D NMR experiments. Figure A.6 describes common 3D NMR experiments and the information that is provided by each. Initial assignments are made using three pairs of complementary experiments: i) HNCACB and CBCA(CO)NH, ii) HNCA and HN(CO)CA, and iii) HNCO and HN(CA)CO. These experiments connect the nuclei of residue *i* to residue *i-1*, allowing

the NMR experimentalists to "walk" down the resonances of the backbone in a sequential manner. The HNCACB and CBCA(CO)NH experiments are paired to define $C_{\alpha(i-1)}$, $C_{\beta(i-1)}$, $H^{N}_{(i)}$, $N_{(i)}$, $C_{\alpha(i)}$, and $C_{\beta(i)}$ resonances. The HNCA and HN(CO)CA are paired to define the $C_{\alpha(i-1)}$, $H^{N}_{(i)}$, $N_{(i)}$, and $C_{\alpha(i)}$ resonances. The HNCO and HN(CA)CO experiments are paired to define the $C'_{(i-1)}$, $H^{N}_{(i)}$, $N_{(i)}$, and $C'_{(i)}$ resonances. These experiments, as well as others that may also be used in sequential backbone assignments are shown below in Figure A.6.

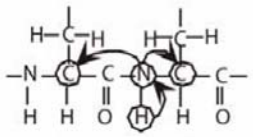
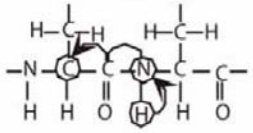
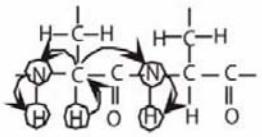
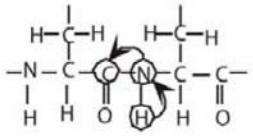
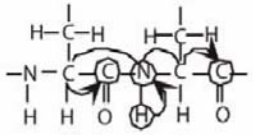
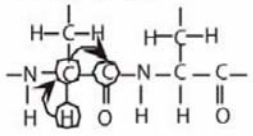
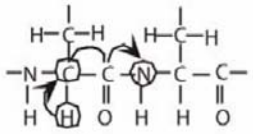
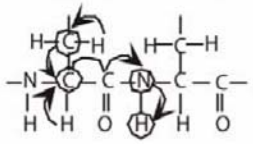
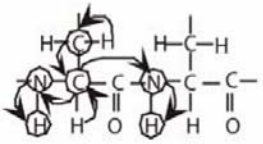
Experiment	Correlations observed	Magnetization transfer	J Couplings
HNCA	$^{1}H_i^N - {}^{15}N_i - {}^{13}C_i^{\alpha}$ $^{1}H_i^N - {}^{15}N_i - {}^{13}C_{i-1}^{\alpha}$		$^{1}J_{NH}$ $^{1}J_{NC\alpha}$ $^{2}J_{NC\alpha}$
HN(CO)CA	$^{1}H_i^N - {}^{15}N_i - {}^{13}C_{i-1}^{\alpha}$		$^{1}J_{NH}$ $^{1}J_{NCO}$ $^{1}J_{C\alpha CO}$
H(CA)NH	$^{1}H_i^{\alpha} - {}^{15}N_i - {}^{1}H_i^N$ $^{1}H_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$		$^{1}J_{C\alpha H\alpha}$ $^{1}J_{NC\alpha}$ $^{2}J_{NC\alpha}$ $^{1}J_{NH}$
HNCO	$^{1}H_i^N - {}^{15}N_i - {}^{13}CO_{i-1}$		$^{1}J_{NH}$ $^{1}J_{NCO}$
HN(CA)CO	$^{1}H_i^N - {}^{15}N_i - {}^{13}CO_i$ $^{1}H_i^N - {}^{15}N_i - {}^{13}CO_{i-1}$		$^{1}J_{NH}$ $^{1}J_{NC\alpha}$ $^{2}J_{NC\alpha}$ $^{1}J_{C\alpha CO}$
HCACO	$^{1}H_i^{\alpha} - {}^{13}C_i^{\alpha} - {}^{13}CO_i$		$^{1}J_{C\alpha H\alpha}$ $^{1}J_{C\alpha CO}$
HCA(CO)N	$^{1}H_i^{\alpha} - {}^{13}C_i^{\alpha} - {}^{15}N_{i+1}$		$^{1}J_{C\alpha H\alpha}$ $^{1}J_{C\alpha CO}$ $^{1}J_{NCO}$
CBCA(CO)NH	$^{13}C_i^{\beta} - {}^{13}C_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$		$^{1}J_{CH}$ $^{1}J_{C\alpha C\beta}$ $^{1}J_{NCO}$ $^{1}J_{C\alpha CO}$ $^{1}J_{NH}$
CBCANH	$^{13}C_i^{\beta}/{}^{13}C_i^{\alpha} - {}^{15}N_i - {}^{1}H_i^N$ $^{13}C_i^{\beta}/{}^{13}C_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$		$^{1}J_{CH}$ $^{1}J_{C\alpha C\beta}$ $^{1}J_{NC}$ $^{2}J_{NC\alpha}$ $^{1}J_{NH}$

Figure A.6. Triple resonance experiments for the assignment of $^{15}N/^{13}C$-labeled proteins.

The resonances for nuclei that are duplicated between the six experiments are averaged to define a single spin system, which is made up of the resonance assignments for $C_{\alpha(i-1)}$, $C_{\beta(i-1)}$, $C'_{(i-1)}$, $H^{N}_{(i)}$, $N_{(i)}$, $C_{\alpha(i)}$, $C_{\beta(i)}$, and $C'_{(i)}$ resonances. Once this is done spin systems are linked based on *i-1* to *i* connectivities, and linked spin systems are assigned to the primary amino acid sequence of the protein being studied. This is facilitated by the fact that certain amino acids have defining characteristics, such as a unique number of atoms or chemical shift profiles that are significantly different from other amino acids. Once all of the resonances are assigned for the backbone, the sidechain carbon and proton must be assigned. This is accomplished with a further set of 3D NMR experiments which include the HCCH-TOCSY and the C(CO)NH.

Types of Structural Restraints

In order to develop a three-dimensional NMR solution structure, structural restraints must be collected. Four types of structural restraints were used in the structure calculations described in this work: i) NOEs, ii) dihedral angles, iii) hydrogen bonds, and iv) residual dipolar couplings.

NOEs. NOEs are considered the single most important type of structural restraint in terms of calculating a high-resolution NMR-based three-dimensional structure. As described above, NOESY experiments provide through-space correlations, and the intensity of these

correlations can provide information about the distance between the two nuclei. In this work, the isolated spin pair approximation was used to calculate distances of NOEs between various nuclei based on known reference distances using the following equation:

$$r_i = r_{ref} (S_{ref}/S_i)^{1/6} \quad \text{(Equation A6)}$$

where r_{ref} is the reference distance, S_{ref} is the intensity of the NOE cross peak corresponding to the reference distance, S_i is the intensity of the NOE cross peak of interest, and r_i is the distance of interest.

Dihedral angles. There are two methods for determining the dihedral angle constraints in a protein. They can be determined directly through an experiment such as the HNHA which measures the scalar coupling constant ^3J which allows for the determination of phi and psi angles by the Karplus equation:

$$^3J = A \cos^2\theta + B \cos\theta + C \quad \text{(Equation A7)}$$

It is also possible to indirectly estimate phi and psi angles using predictive software such as the TALOS program which relies upon a database of residue chemical shift patterns correlated to secondary structure [3]. This is accomplished by inputting the sequence of the protein along with the chemical shift assignments for the backbone H^N, N, C_α, C_β, and C′ nuclei.

Hydrogen bond determination. Hydrogen bonds in a sample are generally collected in one of several ways [4-6]. The most frequently used method involves a hydrogen/deuterium exchange of the sample. A base ^{15}N – ^{1}H correlation spectrum of a sample in 90%:10% (H_2O:D_2O) buffer is collected, which allows for visualization of all amide hydrogens. The sample is then quickly exchanged into a 100% D_2O buffer. The exchange of labile backbone protons for deuterium from the buffer is monitored by collecting a series of ^{15}N – ^{1}H correlation spectra, which will manifest as peaks disappearing in the spectra. The exchange is then generally split into groups that exchange at fast, medium, and slow rates, which allows for the estimation of hydrogen bonds between particular residues when combined with any known information about the secondary or tertiary structure of the protein. It is also possible to utilize experiments that detect the $^{3}J_{NC'}$ couplings across hydrogen bonds. Thus in this case, a 3D HNCO-based experiment is collected, making it possible to unambiguously assign both the hydrogen bond donor and the hydrogen bond acceptor.

Residual Dipolar Couplings. Dipolar coupling based restraints compliment the traditional structural restraints that are strictly local in nature, such as the NOE, J-coupling, hydrogen bond, and dihedral angle [7-8]. However, residual dipolar couplings are not strictly local in nature and can provide a global-based restraint. The dipolar coupling, as described above, is a through-space interaction between two nuclei and is given by the equation below:

$$D_{AB}(\theta,\phi) = A_a^{AB}\{(3\cos^2\theta - 1) + 3/2\ R(\sin^2\theta\cos2\psi)\} \quad \text{(Equation A8)}$$

Where A_a^{AB} and R are the axial and rhombic components, respectively, of the molecular alignment tensor, **A**, in the principle coordinate frame, θ is the angle between the internuclear bond vector and the z-axis of the alignment tensor, ψ is the angle between the projection of the internuclear bond vector onto the x,y plane and the x-axis. A_a^{AB} is given by:

$$A_a^{AB} = -[(\mu_0\hbar)/(16\pi^3)]\, S\gamma_A\gamma_B \langle r_{AB}^{-3}\rangle A_a \quad \text{(Equation A9)}$$

where μ_0 is the permeability in a vacuum, ħ is Planck's constant, S is the generalized order parameter, γ_A and γ_B are the gyromagnetic ratios of the two nuclei, <r> is the time-averaged internuclear distance, and A_a is the axial component of the molecular alignment tensor. Due to rotational diffusion, dipolar couplings are not normally observed in solution NMR spectroscopy ($D_{AB} = 0$). Therefore the sample is dissolved in an anisotropic media or with a co-solute such as Pf1 phage can be introduced to cause partial alignment of the sample so that the θ and ϕ components do not average to zero [9-11].

Structural Calculations

Once all of the chemical shifts have been assigned, NOE restraints, hydrogen bonds, dihedral angle restraints, and RDC restraints are integrated and processed by ARIA and CNS [12-13]. These programs function by incorporating all of the structural restraint information and

producing a final structure based on the available information. This is a highly iterative process in which errors are fixed or removed and restraint information is added until all NMR information is accounted for.

Assessment of quality in NMR structures. Quality assessment of the final structures is determined as a function of how well the structures reflect the experimentally derived constraints and to what degree of precision the structures are determined. Several parameters are generally considered for assessing how well the structures reflect experimental data, which include the number of constraint violations, the number of constraint violations over a certain threshold, the magnitude of violations, energies of the final structures, and the root-mean-squared-deviation (r.m.s. deviation) of the lowest energy structures. High precision will occur if all structures converge to a similar topology. NMR data is inherently "loose" in that NOE data does not have a defined distance associated with the experimental value. Additionally, distance geometry methods randomly assign spatial coordinates to an atom with a range of values. Thus there is an inherent randomness in the NMR distance and dihedral data and starting structures. As a rule of thumb, a high-resolution structure is defined as having more than 15 restraints per residue, no consistent violations present in all structures, no NOE distance violations greater than 0.5 Å, no dihedral angle violations greater than 5^{o}, and a r.m.s. deviation from the mean structure less than 0.8 Å. Below is schematic representation of the steps necessary to obtain a high-confidence and high-resolution NMR solution structure.

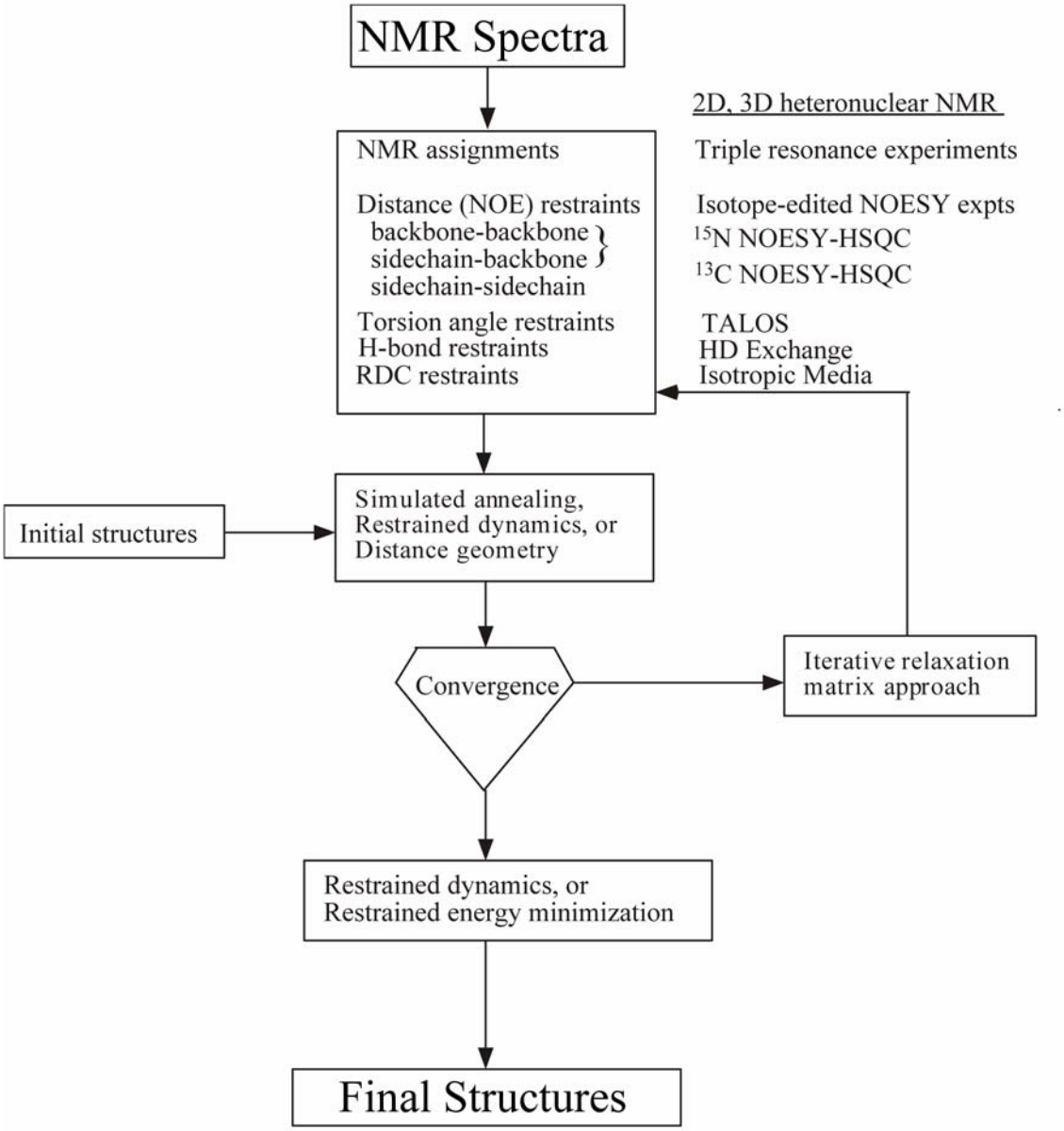


Figure A.7. Schematic representation of the overall process of structure determination by NMR spectroscopy

Dynamic Analysis of Proteins

One of the features that sets NMR structural studies apart from other methodologies is the ability to experimentally determine protein dynamics by measuring relaxation rates for nuclei in the protein [1,14]. Relaxation refers to the process by which a nuclear spin that has been perturbed to a high energy state makes the transition back to the low energy equilibrium state. Accurate measurement of relaxation rates allows for the determination of the variety of timescales over which the motion of a nucleus takes place, ranging from the picosecond to millisecond timescales to seconds or longer. It is thought that these motions are correlated to the functionality of the protein, and understanding these dynamics are as critical to understanding protein function as is determining the high-resolution three-dimensional structure.

The dynamic characteristics of a protein are most commonly described through the analysis of the amide ^{15}N nucleus from measured ^{15}N relaxation rate constants ($1/T_1$ and $1/T_2$) and steady-state NOEs. This relaxation process is dominated by dipolar interactions with the directly bound amide proton and chemical shift anisotropy of the N-H bond vector:

$$R_1 = 1/T_1 = 1/T_1^{DD} + 1/T_1^{CSA} \quad \text{(Equation A10)}$$

$$R_2 = 1/T_2 = 1/T_2^{DD} + 1/T_1^{CSA} + R_{ex} \quad \text{(Equation A11)}$$

where DD is the dipole-dipole interactions, CSA is the chemical shift anisotropy, and R_{ex} is representative of processes such as conformational exchange that contribute to transverse

relaxation properties. Relaxation properties are due to fluxuations of the N-H bond vector in length and orientation relative to the external magnetic field. Such fluxuations are caused by random protein tumbling in solution, local protein motion, and hydrogen chemical exchange.

NMR relaxation parameters can be expressed in terms of a spectral density function, $J(\omega)$, that represents the distribution of re-orientational frequencies expressed by the N-H bond vector. The relaxation rates are described by [15]:

$$R_1 = (d^2/4)[J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H+\omega_N)] + C^2J(\omega_N) \quad \text{(Equation A12)}$$

$$R_2 = (d^2/8)[4J(0) + J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H+\omega_N+]+$$

$$(c^2/)[4J(0) + 3J(\omega_N)] \quad \text{(Equation A13)}$$

$$\sigma_{NH} = (d^2/4)[6J(\omega_H+\omega_N) - J(\omega_H-\omega_N)] \quad \text{(Equation A14)}$$

where d, the strength of the dipolar interaction between the amide nitrogen and proton

$$d = (\mu_0\hbar\gamma N\gamma H/8\pi^2r^3) \quad \text{(Equation A15)}$$

where μ_0 is the permeability of free space, $\hbar$ is Planck's constant, γN and γH are the gyromagnetic ratios of ^1H and ^{15}N, and r is the N-H bond length.

c, the strength of the chemical shift anisotropy,

$$c = \Delta\sigma\omega_N/(3)^{1/2} \quad \text{(Equation A16)}$$

The model-free formalism has been most widely applied to biomolecules and simplifies the analysis with the following [16,17]:

$$J(\omega) = 2/5[\{S^2\tau_m/(1+\omega^2\tau_m^2)\} + \{(1-S^2)\tau/(1+\omega^2\tau^2)\}] \quad \text{(Equation A17)}$$

where τ_m is the isotropic correlation time (and $\tau_m = 1/6D$; where D is the diffusion tensor), S^2 is the generalized order parameter and $\tau = \tau_e\tau_m/(\tau_e+\tau_m)$ where τe is the effective (or local) correlation time. This simplification has fewer parameters (τ_e,τ_m and S^2) that can be fitted by three experiments: inversion recovery (for T_1) [18], CPMG (for T_2) [19], and the steady state NOE [20]. The S^2 order parameter can be interpreted as the N-H vector precessing in a cone that represents the orientational distribution in which a value of 1 is a fixed vector and a value of 0 has isotropic reorientational capacity.

Extraction of relaxation rates from T_1 and T_2 data s completed by fitting the relaxation curves to single exponential decay functions [20,21]:

$$I(t) = I_\infty - [I_\infty - I_0]\exp^{(-t/T1)} \quad \text{(Equation A18)}$$

and

$$I(t) = I_0\exp^{(-t/T2)} \quad \text{(Equation A19)}$$

where I is the resonance intensity at time "t", the initial intensity "0", or the final intensity "∞." The CSA contribution from the amide proton to the amide nitrogen from the steady-state NOE experiment is determined from:

$$NOE = (I_{sat} - I_{equil})/I_{equil} \quad \text{(Equation A20)}$$

where I_{sat} is the NOE intensity at full proton resonance saturation or I_{equil} at equilibrium.

Monte Carlo simulations for the decay curves are performed based on input parameters acquired from these experimentally derived data. These simulations include the single S^2

order parameter to account for internal motions contributing to the characteristic relaxation curve. A statistical analysis is completed and a χ^2 variable is obtained. Subsequent relaxation simulations include additional parameters to determine if the relaxation decays are better described as determined by the comparison of χ^2 and an F-test [22]. The models in this approach are:

Model 1: S^2 only

Model 2: S^2, τe

Model 3: S^2, R_{ex}

Model 4: S^2, τ_e, R_{ex}

Model 5: S^2, τ_e, S_f^2

The additional parameters account for motion on the following timescales: τ_e = picosecond – nanosecond, R_{ex} = microsecond – millisecond, and S_f^2 = motions < 10 picoseconds.

Appendix B

Mass Spectrometry

Mass spectrometry is a powerful analytical technique by which the mass of a molecule can be determined with a very high degree of sensitivity. It is essentially a technique for "weighing" molecules, based on the motion of a charged particle, an ion, in an electric or magnetic field. The mass to charge ratio (*m/z*) of the ion affects this motion. Once formed, the ions are electrostatically drawn into an analyzer where they can be separated based on their *m/z* ratios and then detected. An analogy can be drawn between the processes that occur in a mass spectrometer and a prism. In the prism light it separated into its component wavelengths, which can then be visualized. In the mass spectrometer, ions are separated in the mass analyzer and then detected by an ion detector.

The mass spectrometer can be broken into four basic components: i) a sample inlet, ii) an ionization source, iii) a mass analyzer, and iv) an ion detector. Though some instruments will combine certain components, all molecules will undergo the same processes of being introduced into the instrument through an inlet, converted to ions by the ionization source, electrostatically drawn into the analyzer for separation according the their *m/z*, and then finally detected, converting ion energy into an electrical signal that is transmitted to a computer. Several issues arise from coupling the ion source. The mass spectrometer is kept

a very high vacuum (~10^{-6} torr) to prevent collisions between ion molecules and other gaseous molecules or atoms while ionization will, in some cases, take place at atmospheric pressures (760 torr). Thus a sample must be introduced in such a way that the vacuum inside the instrument remains relatively unchanged. This can be accomplished by one of several ways, direct insertion with a probe or plate, or direct infusion or injection into an ionization source.

Once a sample is introduced, it can be ionized by several types of sources. There are two main types of ionization sources: i) liquid phase ion sources which includes electrospray (ESI), thermospray and atmospheric pressure chemical ionization, and ii) solid state ion sources which include matrix-assisted laser desorption (MALDI) secondary ion mass spectrometry, field desorption, and plasma desorption sources. The work described in this thesis utilized a variant of ESI, which produces gaseous ionized molecules directly from a liquid solution by creating a fine spray of highly charged droplets in the presence of an electric field. This variant is known as nanoESI, in which the flow rate if analyte is on the order of tens to hundreds of nanoliters per minute. A liquid containing the analyte is pumped at low flow rates through a stainless steel capillary where a voltage is applied, which produces an electrical gradient on the fluid, separating charges at the surface. This forces the liquid to emerge from the capillary as a Taylor cone. This cone forms a filament until the liquid reaches the Rayleigh limit for a cylindrical system:

$$q^2 = 6\pi^2\varepsilon\sigma r l^2 \quad \text{(Equation B1)}$$

where ε is the permittivity of the medium surrounding the cylinder, σ is the surface tension, r is the radius of the cylinder and l is the length of the cylinder. Once this limit has been reached highly charged droplets leave the filament, attracted towards the high opposite voltage at the entrance of the mass spectrometer. Dry gas and/or heat are applied to the droplets at atmospheric pressure causing the solvent to evaporate from each droplet. As the droplets rapidly decrease in size due to solvent evaporation, the charge density on the surface will increase. Eventually the Rayleigh limit for a spherical system will be met:

$$q^2 = 8\pi^2\varepsilon\sigma r^3 \quad \text{(Equation B2)}$$

At this point the mutual Coulombic repulsion between like charges on the surface of the droplet become so great that it exceeds the forces of surface tension, and the droplet will explode into smaller droplets, ultimately releasing ions that are attracted through electrostatic lenses leading to the vacuum of the mass analyzer of the instrument.

In its most simplistic form, a mass analyzer separates gas phase ions with respect to their *m/z* ratios, where the charge is produced by the addition or loss of one or more protons, cations, anions, or electrons. This charge allows the molecule to be affected to electric fields, thereby allowing the measurement of mass. Though all mass spectrometers rely on a mass analyzer, not all mass analyzers operate on the same principles; some may separate ions in space, while others separate ions in time. The performance of a mass analyzer can be defined by three

main characteristics: accuracy, resolution, and mass range. *Accuracy* is defined as the ability with which the mass analyzer can provide *m/z* information. Thus an instrument with an accuracy of 0.01% can provide information on a 1000 Da molecule to ± 0.1 Da. Accuracy may also be described using part per million (ppm) terminology, where the above accuracy of ± 0.1 Da corresponds to ± 100 ppm. *Resolution* is the ability of a mass spectrometer to distinguish between ions of different *m/z* ratios. Resolution can be defined by the following equation:

$$\text{Resolution} = m/\Delta m \quad \text{(Equation B3)}$$

where m corresponds to *m/z* and Δm corresponds to the full width at half maximum (FWHM). Two peaks can be said to be resolved if the valley between the two peaks is between 10 – 50% of the weaker peak intensity. This is illustrated below.

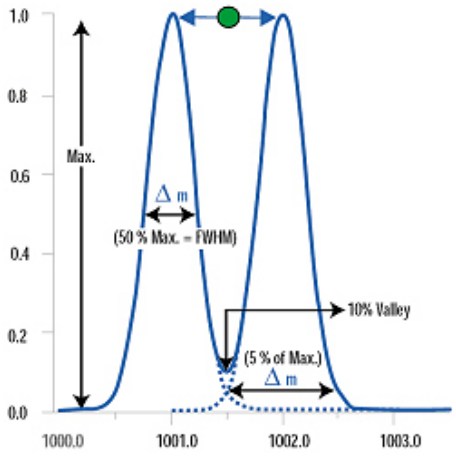


Figure B.1. Example of resolution in a mass spectrum. Note Δm is defined as the width of a peak at 50% maximal height

The final property that determines the performance of a mass analyzer is the *mass range*, which is dependent on the type of mass analyzer utilized. Put simply, the mass range is the *m/z* range of the mass analyzer.

Given the number of possible ionization sources, there is an equally wide variety of mass analyzers with varying performance criteria. Several of the available mass analyzers are quadrupole, ion traps, time-of-flight, and ion cyclotron resonance. The type of mass analyzer utilized for the work presented in this thesis was a hybrid quadrupole time-of-flight (Q-TOF) mass analyzer, which will be discussed below.

The Q-TOF is a coupled mass spectrometric instrument in which a quadrupole (Q) mass analyzer is coupled to a downstream time-of-flight (TOF) mass analyzer. The quadrupole mass analyzers are an arrangement of four electronic rods connected in parallel to a radio frequency generator and a DC potential. This type of mass analyzer works by varying the RF field that is generated. At specific RF fields, only ions of a specific *m/z* can pass through the quadrupoles, while ions of other *m/z* ratios are deflected to the side of the quadrupole. The time-of-flight mass analyzer is the simplest mass analyzer. It is based on accelerating a group of ions through an accelerating potential to impart the same energy to all of the ions. Because the ions have the same energy, lighter ions will reach the detector first because of their greater velocity, while heavier ions will take longer due to their lower velocity. Thus ions are separated by the time required to reach the detector. The theory underlying time-of-

flight may be described using the equations in the following section.

The potential energy of an ion in an electric field is related to its charge and the strength of the electric field:

$$E_p = zU \quad \text{(Equation B4)}$$

where E_p is the potential energy, z is the charge of the ion, and U is the electric potential difference (voltage). When accelerated into the TOF mass analyzer by the voltage, the potential energy is converted into kinetic energy (E_k):

$$E_k = \tfrac{1}{2}\, mv^2 \quad \text{(Equation B5)}$$

Where m is the mass of the ion and v is its velocity. Thus, the above equations can be reordered as:

$$zU = \tfrac{1}{2}\, mv^2 \quad \text{(Equation B6)}$$

Because

$$v = d/t \quad \text{(Equation B7)}$$

where d is the distance of the TOF mass analyzer and t is the time necessary to travel this distance. This allows the following substitutions:

$$zU = \tfrac{1}{2}\, m\, (d/t)^2 \quad \text{(Equation B8)}$$

which can be rearranged as:

$$m/z = (t^2 2U)/(d^2) \quad \text{(Equation B9)}$$

When a quadrupole mass analyzer and a TOF mass analyzer are utilized in tandem, the result is a Q-TOF, a schematic of which is depicted below in Figure B.2.

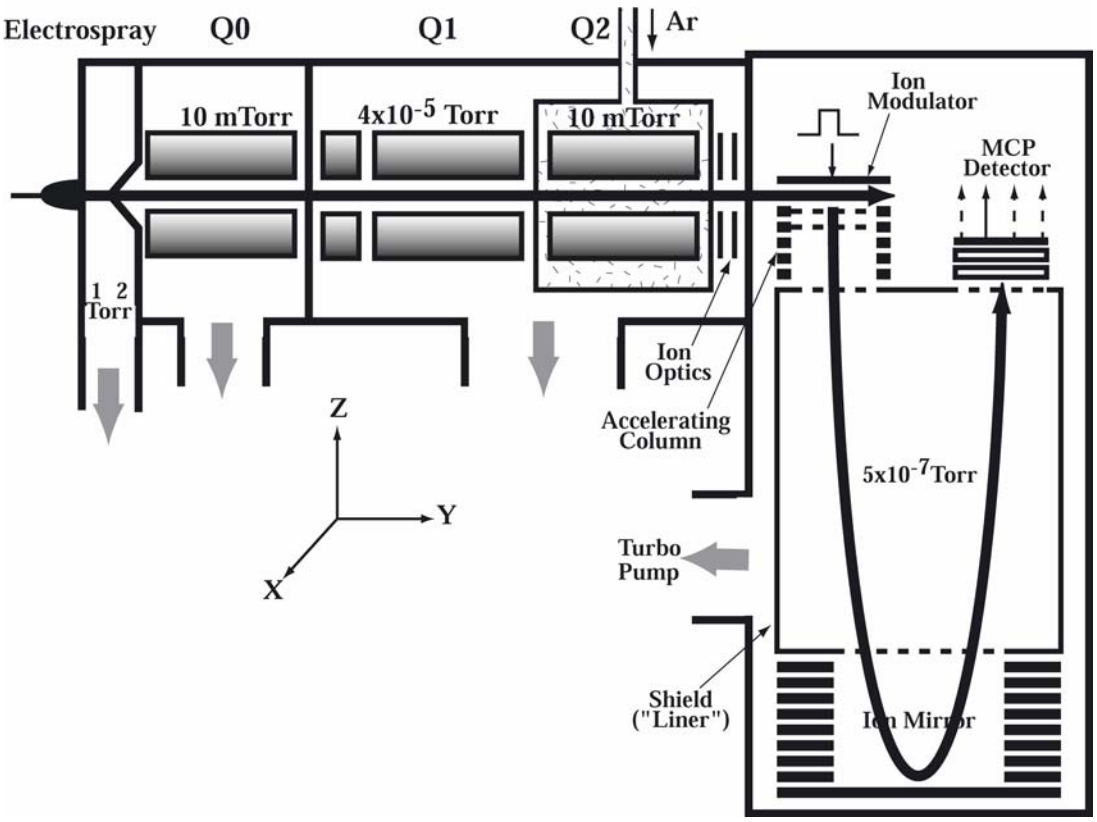


Figure B.2. Schematic diagram of a tandem Q-TOF mass spectrometer.

This instrument consists of three quadrupole (Q0, Q1, Q2) connected in series to a TOF analyzer. Ions generated by the ionization source are directed into Q1 by Q0 which acts as a RF ion guide, focusing the ions as they enter the instrument. By scanning RF frequencies in Q1, it is possible to select ions of specific *m/z* ratio which then enter the TOF analyzer. Is it further possible to carry out tandem mass spectrometry, or MS/MS on this instrument. Q1 is operated in mass filter mode to select the ion of interest, termed the parent ion, which is then

accelerated into the collision cell Q2, where it undergoes collision-induced dissociation (CID) by collision with neutral gas molecules (usually argon or nitrogen). The resulting ion fragments, termed daughter ions, are then introduced into the TOF analyzer. Below is an example mass spectrum from a Q-TOF instrument (Figure B.3).

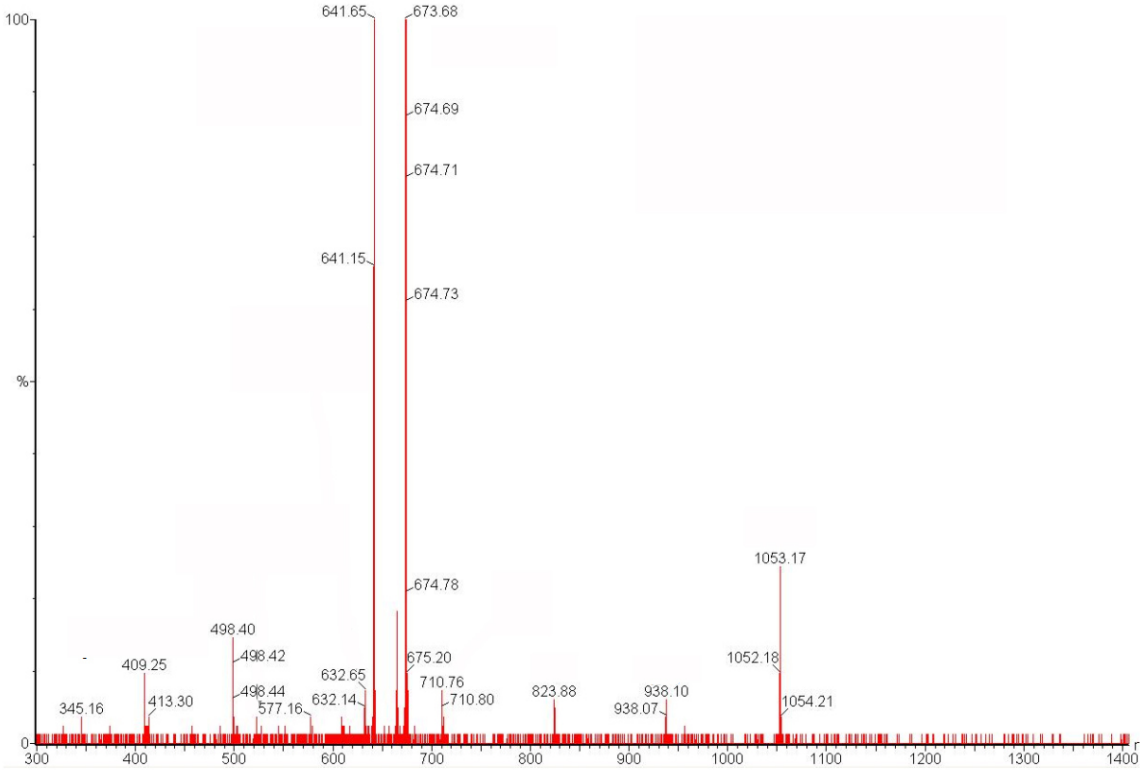


Figure B.3. Example mass spectrum

Once *m/z* values have been obtained, it is necessary to calculate the actual mass of the ions under consideration. Below is an example of how mass calculations are completed:

$$m/z = (MW + nH^{+}) / n \quad \text{(Equation B10)}$$

where *m/z* is the mass to charge ratio, MW is the molecular weight of the ion, n is the integer number of charges on the ion, and H is the mass of a proton (1.008 Da).

Appendix C

Circular Dichroism

Circlar Dichroism (CD) spectroscopy measures the differences in the absorption of left and right-handed circularly polarized light which arise due to structural asymmetry in optically active material. A CD spectropolarimeter is able to measure accurately in the far UV at wavelengths down to 190 – 170nm. In addition, the differences in left and right-handed absorbance A(l) – A(r) is very small (usually in the range of 0.0001), corresponding to an ellipticity of a few $1/100^{th}$'s of a degree. CD spectra for the distinct types of secondary structure present in proteins are different, and thus the analysis of the CD spectrum of a biomacromolecules can yield valuable information about the secondary structure.

Linear polarized light can be viewed as a superposition of opposite circularly polarized light of equal amplitude and phase. A projection of the combined amplitudes perpendicular to the propagation direction yields a line. When this light passes through an optically active sample with a different absorbance (A) for the two components, the amplitude of the stronger absorbed component will be smaller than that of the less absorbed component. The consequence of this is that a projection of the resulting amplitude will now yield an ellipse instead of a line. The occurrence of this ellipticity is called Circular Dichroism.

As stated above, the difference in measured absorption is very small, but it can be determined to a high degree of accuracy. The raw data obtained by CD represents the ellipticity of the

sample in radians:

$$\theta_r = (2.303)/4 * (A_L - A_R) * [rad] \quad \text{(Equation C1)}$$

which can easily be converted into degrees:

$$\theta_d = (2.303)/4 * (A_L - A_R) * [deg] \quad \text{(Equation C2)}$$

In order to compare ellipticity values, they must be normalized. The unit most commonly used in protein and peptide Circular Dichroism spectroscopy is the mean molar ellipticity per residue. In order to accomplish this, the path length (l), concentration (c), molecular weight (M), and the number of residues (n_r):

$$\theta_{mr} = \theta_d * [M/(c*l*n_r)] \quad \text{(Equation C3)}$$

In decimol units, this is:

$$\theta_{mrd} = (\theta_{mr}/10) [deg*(g/dmol)*(cm^3/g)*(1/cm)*(1/residue)] \quad \text{(Equation C4)}$$

which reduces to:

$$\theta_{mrd} = (\theta_{mr}/10) [(deg*cm^3)/(dmol*residue)] \quad \text{(Equation C5)}$$

The values for mean molar ellipticity per residue are usually in the 10,000's.

Each of the three basic secondary structure elements of a polypeptide chain (helix, sheet, and random coil) show a characteristic CD spectrum. A protein composed of these secondary structure elements should therefore display a spectrum that can be deconvoluted into the contributions from the three elements. Example CD spectra is seen below (Figure C.1):

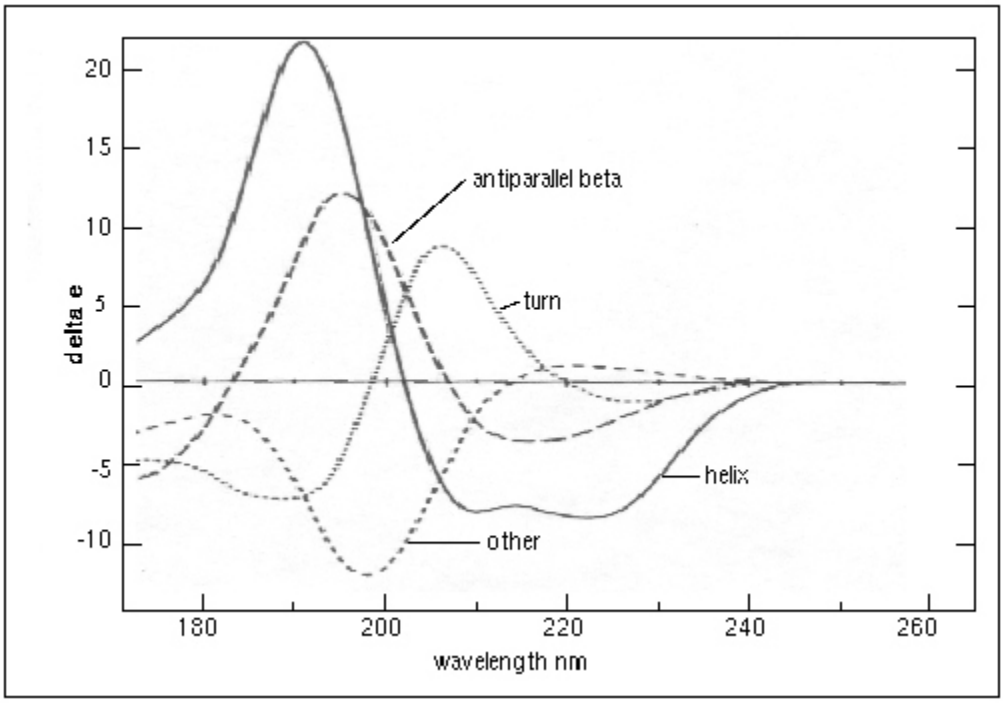


Figure C.1. Example CD spectra. CD spectra for an a-helix (multiple minimum at 208 nm and 220 nm), b-sheet (minimum at 222 nm), and random coil (minimum at 201 nm) of "pure" secondary structures.

Appendix D

Protein Expression and Purification

Construction, Expression, and Purification of the Bacillus subtilis SpoVTN pET-21b(+) expression vector.

Construction of the SpoVT pET-21b(+) expression vector

The SpoVT vector template was provided by Dr. Mark Strauch (The University of Maryland at Baltimore, Baltimore, MD). The gene sequence corresponding to full-length SpoVT was amplified by PCR using the following oligonucleotide primer pair:

SpoVT-Forward:

5'- GTG AGC GCA TAT GAA AGC AAC CGG -3'

SpoVT-Reverse:

5'- CGG GAT CCC TAC TGT TCC ATT TGA CGA -3'

These primers were purchased from IDT-DNA and were designed to incorporate the NdeI restriction endonuclease cleavage site in the forward primer and the BamHI restriction endonuclease cleavage site in the reverse primer. Following PCR amplification and restriction digestion, the cleaved PCR product was ligated into the pET-21b(+) protein expression vector supplied by Novagen. This expression vector places the gene sequence

under an isopropyl-β-D-thiogalactopyranoside (IPTG) inducible promoter and confers ampicillin resistance upon the expression host. This expression vector was introduced into *Esherichia coli* BL21(DE3) cells and tested for the IPTG-inducible overexpression of the target protein. All cells transformed with the SpoVT pET-21b(+) expression vector produced a protein of approximately 19.7 kDa.

Construction of the SpoVTN pET-21b(+) expression vector

In order to express the protein sequence corresponding to the N-terminal 55 residues of SpoVT (SpoVTN), the SpoVT pET-21b(+) expression vector was used as the template for a PCR based mutagenesis reaction according to the protocol provided by the Stratagene QuikChange Site-Directed Mutagenesis Kit. The following oligonucleotide primers purchased from IDT-DNA were designed to incorporate the TAA stop codon at residue 56 in the gene sequence corresponding to full length SpoVT:

SpoVTN Stop Forward:

5'- GCA TAC TCC TTT GCA AAG TCT TAA AGC TCA CTG ATC GGA GAG -3'

SpoVTN Stop Reverse:

5'- CTC TCC GAT CAG TGA GCT TTA AGA CTT TGC AAA GGA GTA TGC -3'

The result of this mutagenesis of the SpoVT pET-21b(+) expression vector is the SpoVTN pET-21b(+) expression vector. When transformed into *E. coli* BL21(DE3) cells, this expression vector produces a protein of approximately 6.4 kDa.

Expression and purification of SpoVTN

SpoVTN pET-21b(+) plasmid DNA isolated using the QIAprep Spin Miniprep Kit from QIAGEN was transformed into BL21(DE3) competent cells from Novagen. Cells from the transformation were plated on 100 μg/mL LB/ampicillin agar plates and placed in a 32^{o}C incubator overnight. The following morning a single colony was used to inoculate 5 mL of LB (w/ 100 μg/mL ampicillin). Once the culture became turbid, 5 mL were used to inoculate a 50 mL LB overnight (w/ 100 μg/mL ampicillin). The following morning 30 mL of the overnight were used to inoculate 1 L of LB media containing 100 μg/mL of ampicillin. The culture was grown at 34^{o}C and 180 rpm to an optical density (OD) A_{600} of 0.950 and protein expression was induced with the addition of IPTG to a final concentration of 1 mM. Growth was allowed to continue for 2 hours before pelleting via centrifugation. Cell pellets were then resuspended on ice in 10 mM Tris-HCl (pH 7.9 at room temperature), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3, 0.25 μM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), and 0.01% Triton X-100. Resuspended cells were then lysed by sonication in 10 cycles of 2 minute bursts/3 minutes rest. Lysed cell suspensions were then centrifuged for 25 minutes at 17,500 rpm. Solid ammonium sulfate was then slowly added to the resulting supernatant to a final concentration of 35% and then allowed to sit for 30 minutes. This mixture was then centrifuged again for 25 minutes at 17,500 rpm. The resulting pellet and supernatant were then checked for the presence of SpoVTN by 12% Tricine gel electrophoresis. As SpoVTN remains soluble at 35% ammonium sulfate, the pellet was discarded and the supernatant was then loaded onto a Phenyl Sepharose (Sigma) column. The column is then washed with 100 mL of the above buffer with 35% ammonium

sulfate. Protein is eluted from the column using a 35% - 0% ammonium sulfate gradient. SpoVTN elutes in approximately the middle of this gradient. Fractions containing SpoVTN were pooled and dialyzed three times against 4L of 10 mM Tris-HCl (pH 7.9), 10 mM KCl, 1 mM EDTA, 1 mM DTT, and 0.02% NaN_3. Dialyzed protein was then loaded onto a Q-Sepharose column (GE Healthcare), the column was washed with 100 mL of the above buffer and eluted using a 10 mM – 500 mM KCl gradient. Again, fractions containing SpoVTN were pooled and dialyzed once against the aforementioned buffer. The protein was then loaded onto a Heparin Agarose column (Sigma), washed with 100 mL of the above buffer, and eluted using a 250 mM – 500 mM KCl gradient. Fractions containing pure SpoVTN were dialyzed against 4L of 10 mM Tris-HCl (pH 7.9), 150 mM KCl, 1 mM EDTA, 1 mM DTT, and 0.02% NaN_3 a total of three times. All purification was performed at 4^{o}C and the presence of SpoVTN was monitored by 12% Tricine gel electrophoresis throughout. Purified protein was concentrated to 1 mM or greater using YM-10 Centriprep centrifugal filtration devices (Millipore).

Expression and Purification of AbrBN and AbhN

Expression and purification of AbrBN

AbrBN pET-24a(+) plasmid DNA isolated using the QIAprep Spin Miniprep Kit from QIAGEN was transformed into BL21(DE3) competent cells from Novagen. Cells from the transformation were plated on 100 μg/mL LB/kanamycin agar plates and placed in a 32^{o}C incubator overnight. The following morning a single colony was used to inoculate 5 mL of LB (w/ 100 μg/mL kanamycin). Once the culture became turbid, 5 mL were used to

inoculate a 50 mL LB overnight (w/ 100 μg/mL kanamycin). The following morning 30 mL of the overnight were used to inoculate 1 L of LB media containing 100 μg/mL of kanamycin. The culture was grown at 37°C and 180 rpm to an optical density (OD) A_{600} of 0.600, at which point the temperature was reduced to 30°C. After one hour protein expression was induced with the addition of IPTG to a final concentration of 1mM. Growth was allowed to continue for 6 hours before pelleting via centrifugation. Cell pellets were then resuspended on ice in 10 mM Tris-HCl (pH 7.9 at room temperature), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3, 0.25 μM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), and 0.01% Triton X-100. Resuspended cells were then lysed by sonication in 10 cycles of 2 minute bursts/3 minutes rest. Lysed cell suspensions were then centrifuged for 25 minutes at 17,500 rpm. Solid ammonium sulfate was then slowly added to the resulting supernatant to a final concentration of 35% and then allowed to sit for 30 minutes. This mixture was then centrifuged again for 25 minutes at 17,500 rpm. The resulting pellet and supernatant were then checked for the presence of AbrBN by 12% Tricine gel electrophoresis. As AbrBN remains soluble at 35% ammonium sulfate, the pellet was discarded and the supernatant was then dialyzed three times against 3L of 10 mM Tris-HCl (pH 7.9 at room temperature), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3. The protein was then loaded onto a Q-Sepharose column (GE Healthcare), the column was washed with 100 mL of the above buffer, and protein was eluted using a 10mM – 500mM KCl gradient. Again, fractions containing AbrBN were pooled and dialyzed once against 3.5 L of the aforementioned buffer. The protein was then loaded onto a Heparin Agarose column (Sigma), the column was washed with 100 mL of the above buffer, and protein was eluted

using a 10 mM – 250 mM KCl gradient. Fractions containing pure AbrBN were dialyzed four times against 4 L 20 mM KH_2PO_4 (pH 5.8), 15 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1 mM $MgCl_2$, 0.02% NaN_3. All purification was performed at 4°C and the presence of AbrBN was monitored by 12% Tricine gel electrophoresis throughout. Purified protein was concentrated to 1 mM or greater using YM-10 Centriprep centrifugal filtration devices (Millipore).

Expression and purification of AbhN

AbhN pET-21b(+) plasmid DNA isolated using the QIAprep Spin Miniprep Kit from QIAGEN was transformed into BL21(DE3) competent cells from Novagen. Cells from the transformation were plated on 100 μg/mL LB/ampicillin agar plates and placed in a 32°C incubator overnight. The following morning a single colony was used to inoculate 5 mL of LB (w/ 100 μg/mL ampicillin). Once the culture became turbid, 5 mL were used to inoculate a 50 mL LB overnight (w/ 100 μg/mL ampicillin). The following morning 30 mL of the overnight were used to inoculate 1 L of LB media containing 100 μg/mL of ampicillin. The culture was grown at 37°C and 180 rpm to an optical density (OD) A_{600} of 0.600, at which point the temperature was reduced to 30°C. After one hour protein expression was induced with the addition of IPTG to a final concentration of 1 mM. Growth was allowed to continue for 6 hours before pelleting via centrifugation. Cell pellets were then resuspended on ice in 10 mM Tris-HCl (pH 7.9 at room temperature), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3, 0.25 μM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride

(AEBSF), and 0.01% Triton X-100. Resuspended cells were then lysed by sonication in 10 cycles of 2 minute bursts/3 minutes rest. Lysed cell suspensions were then centrifuged for 25 minutes at 17,500 rpm. Solid ammonium sulfate was then slowly added to the resulting supernatant to a final concentration of 35% and then allowed to sit for 30 minutes. This mixture was then centrifuged again for 25 minutes at 17,500 rpm. The resulting pellet and supernatant were then checked for the presence of AbhN by 12% Tricine gel electrophoresis. As AbhN remains soluble at 35% ammonium sulfate, the pellet was discarded and the supernatant was then dialyzed three times against 3L of 10 mM Tris-HCl (pH 7.9 at room temperature), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.02% NaN_3. The protein was then loaded onto a Q-Sepharose column (GE Healthcare), the column was washed with 100 mL of the above buffer, and protein was eluted using a 10mM – 500mM KCl gradient. Again, fractions containing AbhN were pooled and dialyzed once against the aforementioned buffer. The protein was then loaded onto a Heparin Agarose column (Sigma), the column was washed with 100 mL of the above buffer, and protein was eluted using a 10 mM – 250 mM KCl gradient. Fractions containing pure AbhN were dialyzed four times against 4L of 20 mM KH_2PO_4 (pH 5.5), 15 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1 mM $MgCl_2$, 0.02% NaN_3. All purification was performed at 4^oC and the presence of AbrBN was monitored by 12% Tricine gel electrophoresis throughout. Purified protein was concentrated to 1 mM or greater using YM-10 Centriprep centrifugal filtration devices (Millipore).

Construction, Expression and Purification of Bacillus subtilis $Spo0F_{His}$ Met81Leu and Met81Gln mutants

Construction of $Spo0F_{His}$ Met81Leu and $Spo0F_{His}$ Met81Gln

The $Spo0F_{His}$ template was provided by Dr. James Hoch. This sequence corresponds to residues 1 – 123 of the 124 residue wild-type Spo0F sequence fused to a non-cleavable C-terminal His-tag. Eight residues (LEHHHHHH) are added to the C-terminus as a result of this tag. This $Spo0F_{His}$ expression vector was used as the template for PCR based mutagenesis reactions performed according to the protocol provided by the Stratagene QuikChange Site-Directed Mutagenesis Kit. The following oligonucleotide primers purchased from IDT-DNA were designed to mutate Met81 of wild-type Spo0F to leucine and glutamine:

Met81Leu

Forward: 5' – ATC CGG GTC ATT ATC CTG ACG GCA TAC GGA GAG – 3'

Reverse: 5' – CTC TCC GTA TGC CGT CAG GAT AAT GAC CCG GAT – 3'

Met81Gln

Forward: 5' – ATC CGG GTC ATT ATC CAG ACG GCA TAC GGA GAG – 3'

Reverse: 5' – CTC TCC GTA TGC CGT CTG GAT AAT GAC CCG GAT – 3'

The results of these mutagenesis reactions are the $Spo0F_{His}$ Met81Leu and $Spo0F_{His}$ Met81Gln expression vectors. When transformed into *E. coli* BL21(DE3) cells, each of these

expression vectors produced a protein of approximately 15.2 kDa.

Expression and Purification of Spo0F$_{His}$ Met81Leu

Spo0F$_{His}$ Met81Leu plasmid DNA isolated using the QIAprep Spin Miniprep Kit from QIAGEN was transformed into BL21(DE3) pLysS competent cells from Novagen. Cells from the transformation were plated on 100 μg/mL LB/ampicillin agar plates and placed in a 32^{o}C incubator overnight. The following morning a single colony was used to inoculate 5 mL of LB (w/ 100 μg/mL ampicillin). Once the culture became turbid, 5 mL were used to inoculate a 50 mL LB overnight (w/ 100 μg/mL ampicillin). The following morning 30 mL of the overnight were used to inoculate 1 L of LB media containing 100 μg/mL of ampicillin. The culture was grown at 30^{o}C and 140 rpm to an optical density (OD) A_{600} of 0.600, at which point protein expression was induced with the addition of IPTG to a final concentration of 0.5 mM. Growth was allowed to continue for 5 hours before pelleting via centrifugation. Cell pellets were then resuspended on ice in 20 mM Tris (pH 7.9 at room temperature), 500 mM NaCl, 0.02% NaN_3, and 0.25 μM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF). Resuspended cells were then lysed by sonication in 10 cycles of 2 minute bursts/3 minutes rest. Lysed cell suspensions were then centrifuged for 25 minutes at 17,500 rpm. The resulting supernatant was loaded onto a Ni-NTA nickel affinity resin (QIAGEN). The column was washed with 150 mL of the above buffer, and protein was eluted using a 0 mM – 250 mM imidazole gradient. Fractions containing Spo0F$_{His}$ Met81Leu were pooled and dialyzed once against 4 L of 25 mM Tris (pH 6.9 at room

temperature), 50 mM KCl, 1 mM EDTA, and 0.02% NaN_3. The inclusion of EDTA at this first stage of dialysis was done to chelate any Ni^{2+} that may have been stripped from the affinity resin during purification. The pure protein was then dialyzed three times against 4 L of 25 mM Tris (pH 6.9 at room temperature), 50 mM KCl, and 0.02% NaN_3. For samples to be used in mass spectroscopic experiments, the protein was dialyzed three times against 4 L of 2 mM $(NH_4)HCO_3$ (pH 6.8 at room temperature). Purification was carried out at room temperature and the presence of $Spo0F_{His}$ Met81Leu was monitored via 15% SDS PAGE. Purified protein was concentrated to the necessary concentrations using YM-10 Centriprep centrifugal filtration devices (Millipore).

Expression and Purification of $Spo0F_{His}$ Met81Gln

$Spo0F_{His}$ Met81Gln plasmid DNA isolated using the QIAprep Spin Miniprep Kit from QIAGEN was transformed into BL21(DE3) pLysS competent cells from Novagen. Cells from the transformation were plated on 100 μg/mL LB/ampicillin agar plates and placed in a 32°C incubator overnight. The following morning a single colony was used to inoculate 5 mL of LB (w/ 100 μg/mL ampicillin). Once the culture became turbid, 5 mL were used to inoculate a 50 mL LB overnight (w/ 100 μg/mL ampicillin). The culture was grown at 34°C and 140 rpm to an optical density (OD) A_{600} of 0.600, at which point the temperature was reduced to 20°C and the rate of shaking was reduced to 120 rpm. An ice water bath was used to cool the temperature of the culture to 20°C. Once the culture reached 20°C, it was allowed to grow for an additional 20 to 30 minutes, at which point protein expression was induced with the addition of IPTG to a final concentration of 0.3 mM. Growth was allowed to

continue for 4 hours before pelleting via centrifugation. Cell pellets were then resuspended on ice in 20 mM Tris (pH 7.9 at room temperature), 500 mM NaCl, 5 mM β-mercaptoethanol (BME), 0.02% NaN_3, and 0.25 μM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF). Resuspended cells were then lysed by sonication in 15 cycles of 2 minute bursts/3 minutes rest. Lysed cell suspensions were then centrifuged for 25 minutes at 17,500 rpm. The resulting supernatant was loaded onto a Ni-NTA nickel affinity resin (QIAGEN). The column was washed with 150 mL of the above buffer, and protein was eluted using a 0 mM – 250 mM imidazole gradient. Fractions containing $Spo0F_{His}$ Met81Gln were pooled and dialyzed once against 4 L of 25 mM Tris (pH 6.9 at room temperature), 50 mM KCl, 5 mM BME, 1 mM EDTA, and 0.02% NaN_3. The inclusion of EDTA at this first stage of dialysis is done to chelate any Ni^{2+} that may have been stripped from the affinity resin during purification. The pure protein is then dialyzed three times against 4 L of 25 mM Tris (pH 6.9 at room temperature), 50 mM KCl, 5 mM BME, and 0.02% NaN_3. For samples to be used in mass spectroscopic experiments, the protein was dialyzed three times against 4 L of 2 mM $(NH_4)HCO_3$ (pH 6.8 at room temperature). Purification was carried out at room temperature and the presence of $Spo0F_{His}$ Met81Gln was monitored via 15% SDS PAGE. Purified protein was concentrated to necessary concentrations using YM-10 Centriprep centrifugal filtration devices (Millipore).

Appendix E

SpoVTN *B. subtilis* Chemical Shift Table

Table E.1. SpoVTN chemical shift values for all assigned residues. Residue number corresponds to the residue in order of appearance of N- to C-terminal. Residue corresponds to the three letter amino acid code. Atom is in agreement with IUPAC identification of atoms. Chemical Shift is given in ppm values.

Residue #	Residue	Atom	Chemical Shift
1	MET	CG	15.937
1	MET	HG2	1.986
2	LYS	CA	55.606
2	LYS	HA	4.54
2	LYS	CB	32.447
2	LYS	HB1	1.865
2	LYS	CG	26.595
2	LYS	HG2	1.489
2	LYS	CD	28.392
2	LYS	HD2	1.683
2	LYS	CE	41.161
2	LYS	HE2	3
2	LYS	C	177.204
3	ALA	N	126.109
3	ALA	HN	8.553
3	ALA	CA	52.218
3	ALA	HA	4.342
3	ALA	CB	19.245
3	ALA	HB1	1.408
3	ALA	C	179.308
4	THR	N	112.465
4	THR	HN	8.078
4	THR	CA	61.583
4	THR	HA	4.42
4	THR	CB	69.46
4	THR	HB	4.204
4	THR	CG2	21.2
4	THR	HG21	1.242
4	THR	C	173.474
5	GLY	N	111.771
5	GLY	HN	9.037
5	GLY	CA	44.687

Table E.1 continued			
5	GLY	HA2	4.058
5	GLY	HA1	3.623
5	GLY	C	174.756
6	ILE	N	121.234
6	ILE	HN	8.287
6	ILE	CA	60.578
6	ILE	HA	4.132
6	ILE	CB	38.785
6	ILE	HB	1.765
6	ILE	CG1	24.463
6	ILE	HG12	1.574
6	ILE	HG11	1.475
6	ILE	CD1	12.335
6	ILE	HD11	0.862
6	ILE	CG2	17.491
6	ILE	HG21	1.043
6	ILE	C	176.395
7	VAL	N	126.096
7	VAL	HN	8.237
7	VAL	CA	60.516
7	VAL	HA	5.002
7	VAL	CB	32.807
7	VAL	HB	1.651
7	VAL	CG2	20.233
7	VAL	HG21	0.263
7	VAL	CG1	21.235
7	VAL	HG11	0.514
7	VAL	C	179.14
8	ARG	N	127.158
8	ARG	HN	9.07
8	ARG	CA	52.551
8	ARG	HA	4.805
8	ARG	CB	34.444
8	ARG	HB2	2.046
8	ARG	HB1	1.613
8	ARG	CG	26.603
8	ARG	HG2	1.936
8	ARG	HG1	1.371
8	ARG	CD	42.841
8	ARG	HD2	3.605
8	ARG	HD1	3.226
8	ARG	C	176.302
9	ARG	N	118.729
9	ARG	HN	8.63
9	ARG	CA	54.299
9	ARG	HA	5.115

Table E.1 continued			
9	ARG	CB	31.277
9	ARG	HB2	1.6
9	ARG	HB1	1.474
9	ARG	CG	27.724
9	ARG	HG2	1.558
9	ARG	HG1	1.292
9	ARG	CD	42.412
9	ARG	HD1	3.19
9	ARG	C	178.367
10	ILE	N	122.469
10	ILE	HN	7.962
10	ILE	CA	58.715
10	ILE	HA	4.541
10	ILE	CB	39.3
10	ILE	HB	1.634
10	ILE	CG1	22.119
10	ILE	HG12	1.325
10	ILE	HG11	1.248
10	ILE	CD1	13.644
10	ILE	HD11	0.702
10	ILE	CG2	16.201
10	ILE	HG21	1.046
10	ILE	C	177.681
11	ASP	N	127.298
11	ASP	HN	8.557
11	ASP	CA	52.44
11	ASP	HA	4.889
11	ASP	CB	40.972
11	ASP	HB2	3.402
11	ASP	HB1	2.832
11	ASP	C	178.929
12	ASP	N	113.147
12	ASP	HN	8.506
12	ASP	CA	55.1
12	ASP	HA	4.054
12	ASP	CB	38.734
12	ASP	HB2	2.454
12	ASP	HB1	2.16
12	ASP	C	178.444
13	LEU	N	120.798
13	LEU	HN	8.268
13	LEU	CA	53.16
13	LEU	HA	4.51
13	LEU	CB	42.04
13	LEU	HB2	1.777
13	LEU	HB1	1.556

Table E.1 continued			
13	LEU	CG	26.504
13	LEU	HG	1.604
13	LEU	CD1	22.893
13	LEU	HD11	0.836
13	LEU	CD2	24.249
13	LEU	HD21	0.829
13	LEU	C	178.487
14	GLY	N	106.617
14	GLY	HN	7.978
14	GLY	CA	44.441
14	GLY	HA2	3.38
14	GLY	HA1	4.168
14	GLY	C	174.619
15	ARG	N	116.828
15	ARG	HN	8.53
15	ARG	CA	54.284
15	ARG	HA	5.202
15	ARG	CB	31.579
15	ARG	HB2	2.222
15	ARG	HB1	1.234
15	ARG	CG	26.559
15	ARG	HG2	1.593
15	ARG	HG1	1.434
15	ARG	CD	42.739
15	ARG	HD2	3.285
15	ARG	HD1	2.7
15	ARG	C	178.314
16	VAL	N	119.216
16	VAL	HN	8.481
16	VAL	CA	58.123
16	VAL	HA	4.494
16	VAL	CB	36.071
16	VAL	HB	1.634
16	VAL	CG2	19.28
16	VAL	HG21	0.723
16	VAL	CG1	20.552
16	VAL	HG11	0.514
16	VAL	C	174.742
17	VAL	N	125.42
17	VAL	HN	7.99
17	VAL	CA	61.551
17	VAL	HA	3.882
17	VAL	CB	32.136
17	VAL	HB	1.871
17	VAL	CG2	21.636
17	VAL	HG21	0.94

Table E.1 continued			
17	VAL	CG1	20.898
17	VAL	HG11	0.829
17	VAL	C	176.487
18	ILE	N	125.876
18	ILE	HN	8.527
18	ILE	CA	56.21
18	ILE	HA	4.332
18	ILE	CB	35.827
18	ILE	HB	1.971
18	ILE	CG1	29.192
18	ILE	HG12	2.141
18	ILE	HG11	1.28
18	ILE	CD1	17.683
18	ILE	HD11	0.927
18	ILE	CG2	21.236
18	ILE	HG21	0.96
18	ILE	C	176.06
19	PRO	CA	62.944
19	PRO	HA	4.549
19	PRO	CB	32.802
19	PRO	HB2	2.689
19	PRO	HB1	1.834
19	PRO	CG	27.694
19	PRO	HG2	1.967
19	PRO	HG1	1.921
19	PRO	CD	50.785
19	PRO	HD2	3.364
19	PRO	HD1	4.437
19	PRO	C	178.97
20	LYS	N	126.618
20	LYS	HN	9.014
20	LYS	CA	59.765
20	LYS	HA	3.756
20	LYS	CB	31.684
20	LYS	HB2	1.955
20	LYS	HB1	1.895
20	LYS	CG	23.878
20	LYS	HG2	1.597
20	LYS	HG1	1.487
20	LYS	CD	28.912
20	LYS	HD2	1.708
20	LYS	HD1	1.165
20	LYS	CE	41.096
20	LYS	HE2	3.035
20	GLU	C	179.313
21	GLU	N	117.455

Table E.1 continued			
21	GLU	HN	9.856
21	GLU	CA	59.771
21	GLU	HA	4.11
21	GLU	CB	28.626
21	GLU	HB2	2.023
21	GLU	CG	36.095
21	GLU	HG2	2.391
21	GLU	HG1	2.326
21	GLU	C	180.774
22	ILE	N	118.288
22	ILE	HN	7.074
22	ILE	CA	63.046
22	ILE	HA	3.849
22	ILE	CB	36.466
22	ILE	HB	1.95
22	ILE	CG1	27.784
22	ILE	HG12	1.556
22	ILE	HG11	1.233
22	ILE	CD1	11.927
22	ILE	HD11	0.799
22	ILE	CG2	17.235
22	ILE	HG21	0.714
22	ILE	C	179.924
23	ARG	N	117.382
23	ARG	HN	7.83
23	ARG	CA	60.379
23	ARG	HA	3.764
23	ARG	CB	28.923
23	ARG	HB2	2.157
23	ARG	HB1	1.718
23	ARG	HG2	1.587
23	ARG	HG1	1.168
23	ARG	CD	43.987
23	ARG	HD2	3.363
23	ARG	HD1	3.157
23	ARG	C	180.11
24	ARG	N	115.785
24	ARG	HN	8.431
24	ARG	CA	58.531
24	ARG	HA	4.183
24	ARG	CB	29.763
24	ARG	HB2	1.924
24	ARG	HB1	2.128
24	ARG	CG	26.173
24	ARG	HG2	1.783
24	ARG	HG1	1.651

Table E.1 continued			
24	ARG	CD	42.396
24	ARG	HD2	3.26
24	ARG	C	181.515
25	THR	N	115.618
25	THR	HN	7.733
25	THR	CA	65.778
25	THR	HA	3.994
25	THR	CB	68.59
25	THR	HB	4.256
25	THR	CG2	21.714
25	THR	HG21	1.328
25	THR	C	177.35
26	LEU	N	117.244
26	LEU	HN	7.796
26	LEU	CA	54.352
26	LEU	HA	4.368
26	LEU	CB	42.33
26	LEU	HB2	1.608
26	LEU	CG	25.107
26	LEU	HG	2.311
26	LEU	CD1	21.914
26	LEU	HD11	0.871
26	LEU	CD2	17.448
26	LEU	HD21	0.774
26	LEU	C	177.163
27	ARG	N	116.574
27	ARG	HN	7.72
27	ARG	CA	56.737
27	ARG	HA	3.872
27	ARG	CB	25.656
27	ARG	HB2	2.085
27	ARG	HB1	1.927
27	ARG	HG2	1.65
27	ARG	CD	42.72
27	ARG	HD2	3.277
27	ARG	C	176.893
28	ILE	N	116.788
28	ILE	HN	8.332
28	ILE	CA	60.527
28	ILE	HA	4.465
28	ILE	CB	39.882
28	ILE	HB	1.562
28	ILE	CG1	23.121
28	ILE	CD1	13.281
28	ILE	HD11	0.76
28	ILE	CG2	18.214

Table E.1 continued			
28	ILE	HG21	0.926
28	ILE	C	176.976
29	ARG	N	128.15
29	ARG	HN	9.187
29	ARG	CA	53.925
29	ARG	HA	4.472
29	ARG	CB	31.701
29	ARG	HB2	1.842
29	ARG	HB1	1.739
29	ARG	CG	25.886
29	ARG	HG2	1.679
29	ARG	HG1	1.608
29	ARG	CD	42.435
29	ARG	HD2	3.221
29	ARG	C	176.397
30	GLU	N	119.127
30	GLU	HN	8.813
30	GLU	CA	58.715
30	GLU	HA	3.642
30	GLU	CB	27.91
30	GLU	HB2	1.952
30	GLU	CG	35.599
30	GLU	HG2	2.156
30	GLU	HG1	2.441
30	GLU	C	178.085
31	GLY	N	113.446
31	GLY	HN	8.365
31	GLY	CA	44.684
31	GLY	HA2	4.401
31	GLY	HA1	3.749
31	GLY	C	175.765
32	ASP	N	121.874
32	ASP	HN	8.288
32	ASP	CA	52.9
32	ASP	HA	4.973
32	ASP	CB	39.908
32	ASP	HB2	3.008
32	ASP	HB1	2.61
32	ASP	C	176.423
33	PRO	CA	61.096
33	PRO	HA	4.774
33	PRO	CB	31.074
33	PRO	HB2	2.119
33	PRO	HB1	1.855
33	PRO	CG	28.577
33	PRO	HG2	1.984

Table E.1 continued			
33	PRO	HG1	1.862
33	PRO	CD	50.276
33	PRO	HD2	4.079
33	PRO	HD1	3.998
33	PRO	C	177.497
34	LEU	N	123.101
34	LEU	HN	9.067
34	LEU	CA	52.842
34	LEU	HA	5.052
34	LEU	CB	44.012
34	LEU	HB2	1.728
34	LEU	HB1	1.237
34	LEU	CG	26.452
34	LEU	HG	1.719
34	LEU	CD1	23.143
34	LEU	HD11	0.786
34	LEU	C	176.051
35	GLU	N	122.921
35	GLU	HN	9.535
35	GLU	CA	54.656
35	GLU	HA	4.218
35	GLU	CB	32.76
35	GLU	HB2	2.18
35	GLU	HB1	1.95
35	GLU	CG	36.096
35	GLU	HG2	2.6
35	GLU	HG1	2.485
35	GLU	C	177.01
36	ILE	N	122.97
36	ILE	HN	8.514
36	ILE	CA	60.838
36	ILE	HA	4.184
36	ILE	CB	38.529
36	ILE	HB	1.888
36	ILE	CG1	26.978
36	ILE	HG12	1.491
36	ILE	HG11	1.212
36	ILE	CD1	13.447
36	ILE	HD11	0.89
36	ILE	CG2	17.458
36	ILE	HG21	0.937
36	ILE	C	177.618
37	PHE	N	125.834
37	PHE	HN	9.392
37	PHE	CA	55.662
37	PHE	HA	5.087

Table E.1 continued			
37	PHE	CB	42.26
37	PHE	HB2	3.094
37	PHE	HB1	3.023
37	PHE	C	175.006
38	VAL	N	116.642
38	VAL	HN	8.519
38	VAL	CA	60.289
38	VAL	HA	5.154
38	VAL	CB	34.258
38	VAL	HB	2.065
38	VAL	CG2	21.315
38	VAL	HG21	1.049
38	VAL	CG1	20.617
38	VAL	HG11	0.772
38	VAL	C	177.654
39	ASP	N	124.423
39	ASP	HN	8.72
39	ASP	CA	51.372
39	ASP	HA	5.014
39	ASP	CB	42.446
39	ASP	HB2	2.827
39	ASP	HB1	3.017
39	ASP	C	178.788
40	ARG	N	117.971
40	ARG	HN	8.601
40	ARG	CA	57.264
40	ARG	HA	4.174
40	ARG	CB	29.007
40	ARG	HB2	1.914
40	ARG	CG	25.857
40	ARG	HG2	1.725
40	ARG	HG1	1.625
40	ARG	CD	42.461
40	ARG	HD2	3.232
40	ARG	C	177.916
41	ASP	N	117.798
41	ASP	HN	8.138
41	ASP	CA	53.578
41	ASP	HA	4.723
41	ASP	CB	39.947
41	ASP	HB2	2.551
41	ASP	HB1	2.875
41	ASP	C	178.048
42	GLY	N	106.809
42	GLY	HN	8.125
42	GLY	CA	46.017

Table E.1 continued			
42	GLY	HA2	4.156
42	GLY	HA1	3.647
42	GLY	C	174.932
43	GLU	N	116.783
43	GLU	HN	7.909
43	GLU	CA	53.603
43	GLU	HA	4.741
43	GLU	CB	29.813
43	GLU	HB2	1.973
43	GLU	HB1	1.875
43	GLU	CG	35.23
43	GLU	HG2	2.135
43	GLU	HG1	2.31
43	GLU	C	176.455
44	VAL	N	120.785
44	VAL	HN	8.635
44	VAL	CA	61.231
44	VAL	HA	4.503
44	VAL	CB	32.334
44	VAL	HB	2.033
44	VAL	CG2	21.031
44	VAL	HG21	0.799
44	VAL	CG1	21.246
44	VAL	HG11	0.961
44	VAL	C	175.366
45	ILE	N	126.64
45	ILE	HN	8.972
45	ILE	CA	59.456
45	ILE	HA	4.889
45	ILE	CB	38.628
45	ILE	HB	1.633
45	ILE	CG1	27.488
45	ILE	HG12	1.077
45	ILE	HG11	1.323
45	ILE	CD1	13.126
45	ILE	HD11	0.764
45	ILE	CG2	17.86
45	ILE	HG21	0.664
45	ILE	C	176.485
46	LEU	N	123.557
46	LEU	HN	9.34
46	LEU	CA	52.946
46	LEU	HA	5.384
46	LEU	CB	42.721
46	LEU	HB2	1.334
46	LEU	HB1	1.725

Table E.1 continued			
46	LEU	CG	25.18
46	LEU	HG	1.638
46	LEU	CD1	18.62
46	LEU	CD2	25.292
46	LEU	HD21	0.719
46	LEU	C	177.393
47	LYS	N	119.462
47	LYS	HN	8.909
47	LYS	CA	53.22
47	LYS	HA	4.959
47	LYS	CB	35.877
47	LYS	HB2	2.039
47	LYS	HB1	1.846
47	LYS	CD	26.326
47	LYS	HD2	1.613
47	LYS	HD1	1.744
47	LYS	CE	40.718
47	LYS	HE2	2.926
47	LYS	HE1	2.836
48	LYS	CA	58.654
48	LYS	HA	4.251
48	LYS	CB	31.479
48	LYS	HB2	2.224
48	LYS	HB1	1.947
48	LYS	CG	24.477
48	LYS	HG2	1.608
48	LYS	HG1	1.463
48	LYS	CD	28.437
48	LYS	HD2	1.825
48	LYS	HD1	1.748
48	LYS	CE	41.197
48	LYS	HE2	3.026
48	LYS	C	178.183
49	TYR	N	125.46
49	TYR	HN	8.189
49	TYR	CA	57.281
49	TYR	HA	4.762
49	TYR	CB	38.697
49	TYR	HB2	2.867
49	TYR	HB1	2.714
49	TYR	C	176.094
50	SER	N	122.04
50	SER	HN	8.045
50	SER	CA	54.28
50	SER	HA	4.635
50	SER	CB	62.402

Table E.1 continued			
50	SER	HB2	3.717
50	SER	HB1	3.601
51	PRO	CA	62.488
51	PRO	HA	4.298
51	PRO	CB	31.121
51	PRO	HB2	1.952
51	PRO	HB1	2.215
51	PRO	CG	26.012
51	PRO	CD	50.211
51	PRO	HD2	3.551
51	PRO	HD1	3.1
51	PRO	C	178.546
52	ILE	N	118.973
52	ILE	HN	7.963
52	ILE	CA	60.234
52	ILE	HA	4.198
52	ILE	CB	37.863
52	ILE	HB	1.917
52	ILE	CG1	26.407
52	ILE	HG12	1.475
52	ILE	HG11	1.21
52	ILE	CD1	12.056
52	ILE	HD11	0.9
52	ILE	CG2	16.688
52	ILE	HG21	0.93
52	ILE	C	177.929
53	SER	N	118.194
53	SER	HN	8.203
53	SER	CA	57.318
53	SER	HA	4.48
53	SER	CB	63.124
53	SER	HB2	3.862
53	SER	C	175.932
54	GLU	N	122.542
54	GLU	HN	8.347
54	GLU	CA	55.56
54	GLU	HA	4.387
54	GLU	CB	29.438
54	GLU	HB2	2.149
54	GLU	HB1	1.931
54	GLU	CG	35.27
54	GLU	HG2	2.307
54	GLU	C	176.98
55	LEU	N	127.38
55	LEU	HN	7.789
55	LEU	CA	54.472

Table E.1 continued			
55	LEU	HA	4.332
55	LEU	CB	29.388
55	LEU	HB2	1.475
55	LEU	HB1	1.298
55	LEU	HG	1.983
55	LEU	HD11	0.809
55	LEU	CD2	9.86
55	LEU	HD21	0.617
55	LEU	C	174.095

References for Appendices

1. Cavanagh, J., Faculty Publication Collection (North Carolina State University), and Wartime Classes Golden Anniversary Endowment, *Protein NMR spectroscopy : principles and practice*. 2nd ed. 2007, Amsterdam ; Boston: Academic Press. xxv, 885 p.

2. Keeler, J., *Understanding NMR spectroscopy*. 2005, Chichester, England ; Hoboken, NJ: Wiley. xv, 459 p.

3. Cornilescu, G., F. Delaglio, and A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, 1999. 13(3): p. 289-302.

4. Wagner, G. and K. Wuthrich, Structural interpretation of the amide proton exchange in the basic pancreatic trypsin inhibitor and related proteins. *J Mol Biol*, 1979. 134(1): p. 75-94.

5. Englander, S.W. and N.R. Kallenbach, Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys*, 1983. 16(4): p. 521-655.

6. Hvidt, A. and S.O. Nielsen, Hydrogen exchange in proteins. *Adv Protein Chem*, 1966. 21: p. 287-386.

7. Lipsitz, R.S. and N. Tjandra, Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct*, 2004. 33: p. 387-413.

8. Clore, G.M., Starich, M.R., and Gronenborn, A.M., Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *Journal of the American Chemistry Society*, 1998. 120: p. 10571-10572.

9. Hansen, M.R., L. Mueller, and A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*, 1998. 5(12): p. 1065-74.

10. Tjandra, N. and A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 1997. 278(5340): p. 1111-4.

11. Chou, J.J., et al., A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR*, 2001. 21(4): p. 377-82.

12. Linge, J.P., S.I. O'Donoghue, and M. Nilges, Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol*, 2001. 339: p. 71-90.

13. Brunger, A.T., et al., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, 1998. 54(Pt 5): p. 905-21.

14. Mittermaier, A. and L.E. Kay, New tools provide new insights in NMR studies of protein dynamics. *Science*, 2006. 312(5771): p. 224-8.

15. Palmer, A.G., 3rd, Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct*, 2001. 30: p. 129-55.

16. Lipari, G.a.S., A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1 Theory and range of validity. *Journal of the American Chemistry Society*, 1982. 104(4546-4559).

17. Lipari, G.a.S., A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemistry Society*, 1982. 104: p. 4559-4570.

18. Vold, R.L., Waugh, J.S., Klein, M.P. and Phelps, D.E., Measurement of spin relaxation in complex systems. *The Journal of Chemical Physics*., 1968. 48: p. 3831-3832.

19. Palmer, A.G., 3rd, C.D. Kroenke, and J.P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*, 2001. 339: p. 204-38.

20. Skelton, N.J., Palmer III, A.G., Akke, M., Kordel, J., Rance, M., and Chazin, W.J., Practical Aspects of Two-Dimensional Proton-Detected 15N spin Relaxation Measurements. *Journal of Magnetic Resonance*, 1993. 102(ser. B.): p. 253-264.

21. Stone, M.J., Fairbrother, W.J., Palmer III, A.G., Reizer, J., Saier Jr., M.H. and Wright, P.J., Backbone Dynamics of the Bacillus subtilis Glucose Permease IIA Domain Determined from 15N NMR Relaxation Measurements. *Biochemistry*., 1992. 31: p. 4394-4406.

22. Mandel, A.M., M. Akke, and A.G. Palmer, 3rd, Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol*, 1995. 246(1): p. 144-63.