MIAMI UNIVERSITY
The Graduate School

Certificate for Approving the Dissertation

We hereby approve the Dissertation

of

Erik Andrew Feldmann

Candidate for the Degree:

Doctor of Philosophy

_____
Michael A. Kennedy, Advisor

_____
Christopher A. Makaroff, Committee Chair

_____
Carole Dabney-Smith, Reader

_____
Michael W. Crowder, Reader

_____
Richard E. Lee, Graduate School Representative

ABSTRACT

BIOPHYSICAL CHARACTERIZATION OF HETEROCYST DIFFERENTIATION REGULATORS, HETR AND PATS, FROM THE CYANOBACTERIUM, ANABAENA SP. STRAIN PCC 7120

AND

STRUCTURAL BIOLOGY OF BACTERIAL PROTEINS FROM THE NORTHEAST STRUCTURAL GENOMICS CONSORTIUM

by Erik A. Feldmann

The filamentous cyanobacterium *Anabaena* (*Nostoc*) sp. strain PCC 7120 has evolved a mechanism for isolating the two incompatible processes of photosynthesis and nitrogen fixation by differentiating specialized cells called heterocysts. The structure and metabolism of heterocysts support a microanaerobic environment allowing for the nitrogenase-dependent fixation of dinitrogen into ammonia to occur with limited inhibition by oxygen. Under conditions of nitrogen starvation, *Anabaena* differentiates heterocysts in regular patterned intervals in approximately one out of every ten cells. The *hetR* and *patS* genes are essential for proper heterocyst differentiation. The gene products, HetR and PatS, are regulators of heterocyst differentiation, but their specific mechanisms of activity are unclear. In the first part of this dissertation, an in depth biophysical characterization of both HetR and PatS is presented. We describe a method for over-expressing and purifying high yields of soluble HetR from *Anabaena* in *Escherichia coli*, have provided the first experimental evidence that the master regulatory

HetR transcription factor is the direct binding target for the PatS inhibitor peptide, have characterized the binding stoichiometry of HetR with DNA and PatS substrates, and have quantified the binding thermodynamics of HetR-PatS complexes. The second part of this dissertation describes a detailed structural biology summary of four different bacterial protein targets of the National Institutes of Health. The first section of this second part includes an uncharacterized protein, Pspto_3016 from *Pseudomonas syringae* which adopts a putative DNA-binding double-wing motif. The second and third sections include two different proteins from the cyanobacterium *Anabaena* sp. PCC 7120: a hypothetical uncharacterized protein, Asl3597, with a novel fold and that exhibits sequence homology to the chlororespiratory reductase family of plant proteins, and another uncharacterized protein, Asr4154, annotated as a subunit of the protochlorophyllide reductase superfamily of enzymes. The fourth and final section includes an uncharacterized protein fragment, the putative mucin-binding domain of the protein LBA1460 from *Lactobacillus acidophilus*. We provide high resolution, three-dimensional structures of each protein, solved by solution NMR spectroscopy, to assist in future characterization of their corresponding biological and biochemical functions.

BIOPHYSICAL CHARACTERIZATION OF HETEROCYST DIFFERENTIATION REGULATORS, HETR AND PATS, FROM THE CYANOBACTERIUM, ANABAENA SP. STRAIN PCC 7120

AND

STRUCTURAL BIOLOGY OF BACTERIAL PROTEINS FROM THE NORTHEAST STRUCTURAL GENOMICS CONSORTIUM

A DISSERTATION

Submitted to the Faculty of

Miami University in partial

fulfillment of the requirements

for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

by

Erik A. Feldmann

Miami University

Oxford, Ohio

2012

Dissertation Director: Michael A. Kennedy

Table of Contents

Chapter 1: Introduction 1
1.1 Heterocyst differentiation in cyanobacteria 2
1.1.1 The emergence of cyanobacteria and the evolution of heterocysts 2
1.1.2 The cyanobacterium, *Anabaena* sp. strain PCC 7120 2
1.2 Regulation of heterocyst pattern formation 4
1.2.1 Initiation of regulatory events 4
1.2.2 HetR, the master regulator of heterocysts 5
1.2.3 PatS, a heterocyst pattern regulator 7
1.3 Structural Genomics 8
1.3.1 The Protein Structure Initiative 8
1.3.2 The Northeast Structural Genomics Consortium 9
1.3.3 Target selection 10
1.3.4 High-throughput cloning, purification, and isotopic labeling of NESG protein targets 10
1.3.5 Multidimensional Nuclear Magnetic Resonance (NMR) spectroscopy 13
1.3.5 Three-dimensional structure calculations 16
1.4 Dissertation goals and specific aims 18
1.6 References 21
Chapter 2: Evidence for direct binding between HetR from *Anabaena* sp. strain PCC 7120 and PatS-5 27
2.1 Abstract 28
2.2 Introduction 28
2.3 Materials and Methods 31
2.3.1 Bacterial strains and growth conditions 32
2.3.2 Plasmid construction for making chromosomal alleles 32
2.3.3 Construction of strains containing mutant alleles 35
2.3.4 *In vivo* PatS-5 sensitivity assays 35
2.3.5 Cloning, overexpression and purification of recombinant soluble HetR 35

2.3.6 DNA binding assays 39
2.3.7 Circular dichroism spectroscopy 39
2.3.8 Site-directed spin labeling 40
2.3.9 Preparation of samples for electron paramagnetic resonance spectroscopy 40
2.3.10 Electron paramagnetic resonance spectroscopy 40
2.3.11 EPR spectral simulations 41
2.3.12 Molecular dynamics simulations 42
2.3.13 Isothermal titration calorimetry 42
2.4 Results 43
2.4.1 Conservative substitutions at HetR residues 250-256 affect heterocyst formation and sensitivity to PatS-5 43
2.4.2 Evidence for direct binding of PatS-5 to HetR 46
2.4.3 Evidence for PatS-5 binding to the dimer form of HetR 52
2.4.4 Characterization of binding of HetR to a 29 bp inverted repeat containing DNA sequence upstream of *hetP* 56
2.4.5 Sensitivity of the HetR-DNA complex to PatS-5 58
2.4.6 Similarity of HetR and HetR-C48A 61
2.4.7 Determination of the stoichiometry, dissociation constant, and thermodynamic parameters for binding of PatS-5 to HetR 65
2.5 Discussion 68
2.6 Acknowledgements 70
2.7 References 70
Chapter 3: Differential binding between PatS C-terminal peptide fragments and HetR from *Anabaena* sp PCC 7120 79
3.1 Abstract 80
3.2 Introduction 80
3.3 Materials and Methods 84
3.3.1 Cloning, expression, and purification of HetR protein and preparation of PatS peptide 84
3.3.2 EPR spectroscopy 84

3.3.3 ITC Calorimetry 85
3.4 Results 86
3.4.1 EMSA knockdown comparison of native PatS peptides 86
3.4.2 Thermodynamics and binding affinities of native and mutant PatS peptides binding to HetR 88
3.4.3 EPR spectroscopy of spin labeled HetR titrated with PatS-6 93
3.5 Discussion 97
3.6 Acknowledgements 99
3.7 References 99
Chapter 4: Structural biology of bacterial proteins from the Northeast Structural Genomics Consortium 104
4.1 Abstract 106
4.2 Introduction to NESG target ID: PsR293 106
4.2.1 Materials and methods 107
4.2.2 Results and Discussion 113
4.3 Introduction to NESG target ID: NsR244 127
4.3.1 Materials and Methods 128
4.3.2 Results and Discussion 137
4.4 Introduction to NESG target ID: NsR143 141
4.4.1 Materials and Methods 141
4.4.2 Results and Discussion 144
4.5 Introduction to NESG target ID: LaR80A 154
4.5.1 Materials and Methods 155
4.5.2 Results and Discussion 158
4.6 References 167
Chapter 5: Conclusions 175
5.1 HetR- and PatS-dependent regulation of heterocyst differentiation in cyanobacteria 176
5.2 Structural Genomics 178
5.3 References 179

List of Tables

Table 2.1 Strains of *Anabaena* sp. strain PCC 7120 and plasmids used to generate mutant *hetR* chromosomal alleles 33

Table 2.2 Primers used to generate chromosomal mutants in *Anabaena* sp. strain PCC 7120..... 34

Table 2.3 Primers, *Escherichia coli* hosts, and expression conditions used to generate recombinant *hetR* mutants 38

Table 2.4 Simulation parameters for room temperature CW X-band EPR spectra of MTSL-labeled HetR 250-256C cysteine mutants and PatS-5 titrations 60

Table 2.5 Thermodynamic quantities of HetR binding to PatS-5 derived from ITC experiments 67

Table 3.1 Thermodynamic quantites of HetR binding to PatS-5, PatS-6, PatS-7, and various C-terminal PatS-6 mutants dervived from ITC experiments 92

Table 3.2 Simulation parameters for room temperature CW X-band EPR spectra of MTSL-labeled HetR 252C titrated with PatS-6..... 96

Table 4.1 Summary of NMR and structural statistics for Pspto_3016 from *Pseudomonas syringae* (PDB ID 2KFP)..... 119

Table 4.2 Summary of crystals, X-ray diffraction data collection, and refinement statistics for Pspto_3016 from *Pseudomonas syringae* (PDB ID 3H9X)..... 121

Table 4.3 Summary of NMR and structural statistics for Asl3597 from *Anabaena* sp. strain PCC 7120 (PDB ID 2KRX)..... 136

Table 4.4 Summary of NMR and structural statistics for Asr4154 from *Anabaena* sp. strain PCC 7120 (PDB ID 2L09)..... 146

Table 4.5 Summary of NMR and structural statistics for the MucBP domain of LBA1460 from *Lactobacillus acidophilus* (PDB ID 2LFI)..... 162

List of Figures

Figure 1.1 Simplified schematic of the chemical components of the nitrogen cycle converted by different microorganisms 3

Figure 1.2 Summarized model including a select group of regulatory control components predicted to be involved in regulating initiation and development of heterocyst differentiation 4

Figure 1.3 X-ray crystal structure of the HetR homodimer from *Fischerella* MV11 (PDB ID 3QOE) shown in a backbone cartoon representation 6

Figure 1.4 Flow chart summarizing the process of high-throughput NMR-based structural genomics at the Miami University Center of Excellence in Structural Biology and Metabonomics for the Northeast Structural Genomics Consortium 12

Figure 1.5 Example of a heteronuclear single quantum coherence experiment (^{1}H-^{15}N-HSQC) optimized for H-N correlations for the protein Pspto_3016 from *Pseudomonas syringae* 14

Figure 1.6 Example 2D projections of NOESY spectra from the MucBP fragment of the protein LBA1460 from *Lactobacillus acidophilus* 15

Figure 1.7 Example ensemble of 20 structures calculated with the program CYANA for the protein Asl3597 from *Nostoc* sp. PCC 7120 17

Figure 2.1 Bar graph and light microscopy images displaying various sensitivities of wild type and mutant hetR strains of *Anabaena* sp. PCC 7120 with PatS-5 45

Figure 2.2 Room temperature CW X-band EPR spectra showing variations in spin label mobility for MTSL-labeled HetR cysteine mutants in the amino acid position of 250-256 49

Figure 2.3 Crystal structure of the HetR homodimer from *Fischerella* (PDB ID 3QOE) with the general location of residues 250-256 depicted by stick representation and MTSL-labeled residues in spheres representation 50

Figure 2.4 CW X-band EPR spectra of MTSL-labeled HetR 252C titrated with PatS-5 51

Figure 2.5 Pulsed DEER EPR data from MTSL-labeled HetR 252C..... 54

Figure 2.6 Backbone cartoon representation of the Fischerella HetR crystal structure (PDB ID 3QOE) modified with a 252C cysteine mutation and covalently-linked MTSL group 55

Figure 2.7 EMSA showing binding of HetR to the 29 base pair inverted repeat-containing fragment of the *hetP* DNA promoter region..... 57

Figure 2.8 EMSA disruption of PatS-5 titrated into complexes of HetR bound to the 29 base pair inverted repeat-containing fragment of the *hetP* DNA promoter region..... 58

Figure 2.9 CW X-band EPR of MTSL-labeled HetR 252C in complex with the 29 base pair inverted repeat-containing fragment of the *hetP* DNA promoter region titrated with PatS-5 and the titration with Poly-G5 59

Figure 2.10 Size-exclusion chromatogram of refolded wild type HetR, soluble-expressed wild type HetR, and C48A HetR..... 62

Figure 2.11 SDS PAGE of wild type and C48A HetR 63

Figure 2.12 Circular dichroism spectra of wild type HetR 64

Figure 2.13 Representative ITC isotherm for HetR binding to PatS-5..... 66

Figure 3.1 EMSA knockdown analysis of PatS C-terminal peptides with the HetR-*hetP*-29-mer DNA complex 87

Figure 3.2 Representative ITC isotherms for native and mutant PatS peptides binding HetR..... 91

Figure 3.3 CW X-band EPR spectra of MTSL-labeled HetR 252C titrated with PatS-6 94

Figure 3.4 Endpoint CW X-band EPR spectrum of MTSL-labeled HetR 252C titrated with PatS-8 95

Figure 4.1 1D ^{15}N relaxation data for [*U*-13C, 15N]-Pspto_3016 from *Pseudomonas syringae* 109

Figure 4.2 2D ^{1}H-^{15}N HSQC spectrum of [*U*-13C, 15N]-Pspto_3016 from *Pseudomonas syringae* 112

Figure 4.3 3D structure of the Pspto_3016 protein from *Pseudomonas syringae* (PDB IDs 2KFP and 3H9X) and representative homologs shown in backbone ribbon and cartoon representations..... 117

Figure 4.4 APBS electrostatic surface potential map, cartoon representations, and ConSurf representations of Pspto_3016 from *Pseudomonas syringae* highlighting specific residues 122

Figure 4.5 Cluster map of the Pfam PF04237 protein family using CLANS 124

Figure 4.6 Sequence alignment of Pspto_3016 from *Pseudomonas syringae* with YjbR from *Escherichia coli* 125

Figure 4.7 Static light scattering data for Asl3597 from *Anabaena* sp. strain PCC 7120 129

Figure 4.8 1D ^{15}N relaxation data for [*U*-13C, 15N]-Asl3597 from *Anabaena* sp. strain PCC 7120 130

Figure 4.9 2D ^{1}H-^{15}N HSQC spectrum of [*U*-13C, 15N]-Asl3597 from *Anabaena* sp. strain PCC 7120..... 133

Figure 4.10 Bar graph of backbone ^{15}N peak linewidths of [*U*-13C, 15N]-Asl3597 for all unique amide H-N peaks in the ^{1}H-^{15}N HSQC spectrum and sausage representation of Asl3597 portraying backbone thickness as a function of ^{15}N linewidth..... 134

Figure 4.11 3D structure of the NMR ensemble, backbone cartoon, ConSurf representation, and APBS electrostatic surface potential map of Asl3597 from *Anabaena* sp. strain PCC 7120 139

Figure 4.12 Multiple sequence alignment of Asl3597 from *Anabaena* sp. PCC 7120 with the top 10 homologous BLAST results 140

Figure 4.13 2D ^{1}H-^{15}N HSQC spectrum of [*U*-13C, 15N]-Asr4154 from *Anabaena* sp. strain PCC 7120 143

Figure 4.14 3D solution NMR structure of the ensemble and backbone cartoon of Asr4154 from *Anabaena* sp. strain PCC 7120 147

Figure 4.15 APBS electrostatic surface potential map of Asr4154 from *Anabaena* sp. strain PCC 7120 149

Figure 4.16 Multiple sequence alignment of Asr4154 from *Anabaena* sp. PCC 7120 and the C-terminal fragment of BchB from *Chlorobium tepidum* 150

Figure 4.17 ConSurf representation of Asr4154 from *Anabaena* sp. strain PCC 7120 151

Figure 4.18 Multiple sequence alignment of Asr4154 from *Anabaena* sp. PCC 7120 and BchB proteins from *Rhodobacter capsulatus* and *Thermosynechococcus elongatus* 153

Figure 4.19 2D ^{1}H-^{15}N HSQC spectrum of [*U*-13C, 15N]-MucBP domain of LBA1460 from *Lactobacillus acidophilus* 157

Figure 4.20 3D solution NMR structure for the MucBP domain of LBA1460 from *Lactobacillus acidophilus* (PDB ID 2LFI) 160

Figure 4.21 ConSurf representation of the MucBP domain of LBA1460 from *Lactobacillus acidophilus* 165

Figure 4.22 Structural alignment of the NMR and crystal structure of the MucBP domain of LBA1460 from *Lactobacillus acidophilus* (PDB IDs 2LFI and 3Q69) and the B2 domain of MubR5 from *Lactobacillus reuteri* (PDB ID 3I57) highlighting specific residues 166

Dedication

This work is dedicated to my family for their unending love and support. To my parents, Joe and Patrice, and to my younger brother, Jeff, who is currently serving abroad as a naval officer onboard a nuclear submarine in the Unites States Navy. Thank you for believing in me.

Acknowledgements

I would like to take this opportunity to thank several people who have contributed during the last five years to the successful completion of this dissertation. I must first send a sincere thanks to my advisor, Dr. Michael Kennedy, for guidance, wisdom, and encouragement in my graduate education. Your passion for the advancement of science at all levels has served as a great example for me to pursue my life-long interests in scientific discovery. I wish you all the best in your future endeavors.

I also would like to thank all the current and former Kennedy Lab graduate students for the bonds that we formed over the years while completing our graduate training in chemistry and maturing together both personally as well as professionally. I would especially like to acknowledge my long-time officemates, Aaron Goodpaster and Lindsey Romick-Rosendale. The nearly four and half years we spent together facing countless joys and challenges, will never be forgotten. Good luck to you both with your careers, and certainly with the wonderful new additions to your families!

To my friends at Miami University – thank you for all the memories. I especially want to acknowledge Dr. Shuisong Ni for all your generous research suggestions, discussions about Chinese and American culture, and your insight on life. I would like to thank Drs. Theresa Ramelot, Yunhuang Yang, and Keyang Ding for the time and effort put forth training me in protein NMR spectroscopy and in the evolving methods of 21st Century high-throughput structural genomics. Also, to Sriram Devanathan, Jayanthi Sanjeevi, Kaushik Chander, Shiva Priya Dharshan, Kuntal De, Debjani Pal, and Sravan Katragadda – thank you all for welcoming me into your families and for teaching me all that you have about Indian culture, everything from cooking to cricket. To Brittany Christian – thank you for being exactly the person I needed in my life, at exactly the right time. I look up to you in so many ways and sincerely appreciate your generosity, compassion, humor, encouragement, witty intellect, and most of all, your friendship.

Lastly, I would like to thank the past and present members of my dissertation committee for their help and direction over the course of my career at Miami: Drs. Christopher Makaroff, Blanton Tolbert, Carole Dabney-Smith, Michael Crowder, Richard Lee, and the late John Hawes, who passed unexpectedly in 2010 and provided with me with a great foundation of scientific mentoring. Thank you John, you are greatly missed.

Chapter 1: Introduction

1.1 Heterocyst differentiation in cyanobacteria

1.1.1 The emergence of cyanobacteria and the evolution of heterocysts

The prehistoric rise of autotrophic microorganisms transformed the planet's early anaerobic atmosphere into one saturated with oxygen. Cyanobacteria are ancient photosynthetic prokaryotes obtaining energy from sunlight and carbon from the atmosphere and first appearing on Earth more than 2 billion years ago (6). Many cyanobacteria are also diazotrophic, capable of "fixing" their own nitrogen source, leading to a wide diversification in their ecological roles from those of high independence to those forming symbiotic relationships with other bacteria and eukaryotes (13). The fixation of nitrogen by cyanobacteria, from the highly unreactive dinitrogen (N_2) species into ammonia (NH_3), is carried out by nitrogenases, a class of oxygen-sensitive, enzyme complexes. Some species of cyanobacteria are capable of separating nitrogenase-dependent nitrogen fixation from oxygenic photosynthesis by compartmentalization into specialized cell types called heterocysts (19, 20). The heterocyst accommodates a microoxic environment via enhancements to its cell wall in the form of an additional hydrophobic glycolipid layer and a heterocyst envelope polysaccharide layer (33, 40). In addition, heterocysts generate less oxygen by downregulating photosynthesis, characterized by the suppression of light-harvesting phycobiliproteins and an absence of photosystem II (18, 23, 46).

1.1.2 The cyanobacterium, *Anabaena* sp. strain PCC 7120

Differentiation of heterocysts is a tightly-regulated process. *Anabaena* (*Nostoc*) sp. strain PCC 7120, whose complete genome is available and fully annotated (30), has been the model cyanobacterium for studying the regulation of heterocyst differentiation and pattern formation. Under normal nitrogen abundant conditions, *Anabaena* grows as vegetative cells along a filamentous chain. When deprived of nitrogen, a pattern begins to take shape where approximately one out of every ten cells differentiates into a heterocyst (46, 47). The heterocyst supplies fixed nitrogen to the neighboring vegetative

cells, and in return receives carbon and reductant to support nitrogen fixation in an interdependent relationship (39). As the filament of *Anaebana* continues to grow under these conditions, additional heterocysts form at the midway point between two preexisting heterocysts, maintaining the semiregular pattern (45).

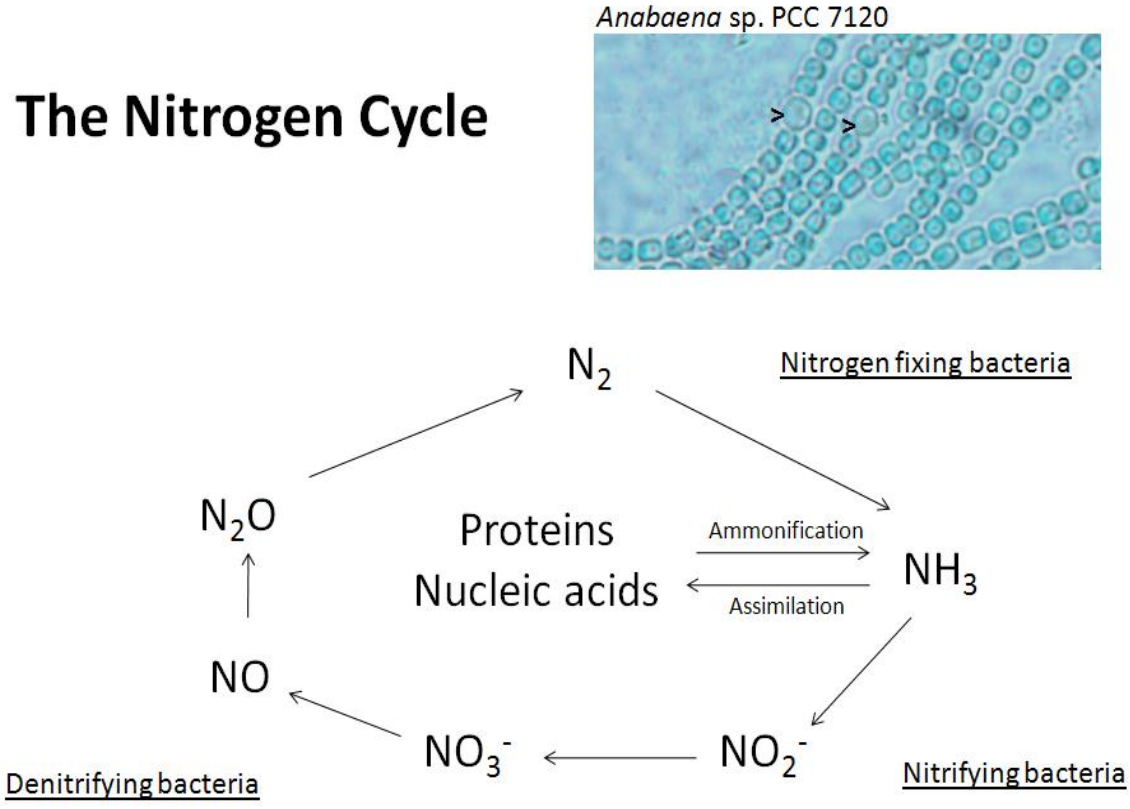

Figure 1.1 Simplified schematic of the chemical components of the nitrogen cycle converted by different microorganisms. *Anabaena* sp. strain PCC 7120 and other nitrogenase-containing diazotrophs belong to the category of nitrogen fixing bacteria indicated above and to the right. Developing heterocyst cells are indicated with carets amidst growing vegetative cells in the filaments.

1.2 Regulation of heterocyst pattern formation

1.2.1 Initiation of regulatory events

Anabaena senses environmental nitrogen deprivation by a metabolic accumulation of 2-oxoglutarate (2-OG, also commonly known as α-ketoglutarate) (35). 2-OG serves as a trigger by stimulating the DNA binding activity of the global nitrogen response protein, NtcA, which is involved in the initiation of heterocysts (25, 34, 43, 44). The gene encoding the NtcA protein, *ntcA*, is required for the early stages of heterocyst development (21). NtcA belongs to the cyclic AMP receptor family of transcription factors, is responsible for activating the expression of many genes required for nitrogen and carbon metabolism (24), and recognizes the consensus nucleotide sequence motif $GTAN_8TAC$ (25). One downstream target of NtcA is NrrA, a member of the OmpR family of transcription factors (17) that has been recently shown to be involved in controlling glycogen catabolism (16). NrrA is a response regulator and activates the expression of the *hetR* gene (15), serving as a regulatory link between NtcA and HetR under the response to combined nitrogen starvation.

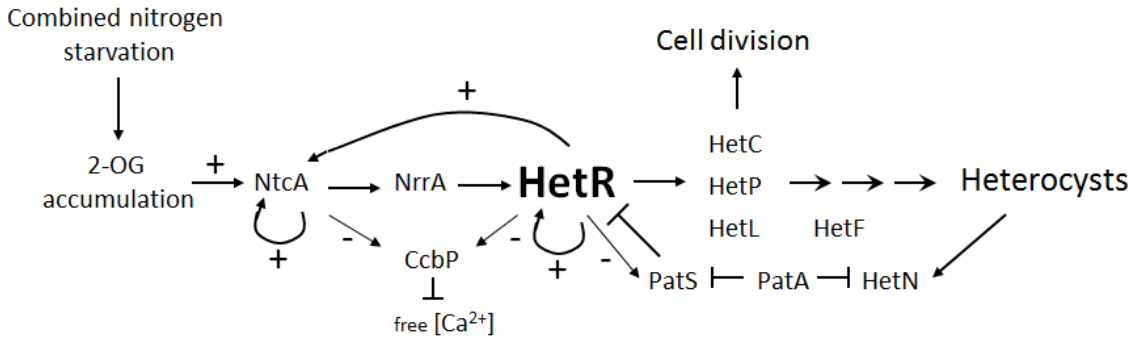

Figure 1.2 Summarized model including a select group of regulatory control components predicted to be involved in regulating initiation and development of heterocyst differentiation. Positive regulation by various components is represented with a "+" sign and negative regulation with a "-" sign.

1.2.2 HetR, the master regulator of heterocysts

HetR is considered a master regulator of heterocyst development. The *hetR* gene is one of the earliest genes to be induced under nitrogen deprivation, is required for the spatial pattern of heterocysts observed along the filaments in *Anabaena*, and also positively regulates its own expression (4). Deletion mutants of *hetR* are incapable of forming heterocysts, while extra copies of *hetR* result in heterocyst differentiation even under nitrogen abundant conditions, and additionally form a multiple contiguous heterocyst phenotype under nitrogen starvation (7). In addition, the expression of *hetR* was shown to have a mutual dependency with *ntcA* expression during heterocyst development (38), suggesting an intimate relationship between the two regulatory elements.

The gene for *hetR* in *Anabaena* is expressed at low levels in vegetative cells and begins to increase approximately two hours after nitrogen step-down and accelerates in developing heterocysts (8). *hetR* encodes for the enigmatic 299-residue protein, HetR, a highly-conserved protein from both heterocyst-forming and non-heterocystous cyanobacteria whose amino acid sequence provides little bioinformatics evidence to its function. HetR is a transcription factor predicted to function as a homodimer and shown experimentally to bind DNA in its own promoter region as well as the promoter regions of the *hepA*, *hetP*, and *patS* genes (27, 29). The only three-dimensional structures of HetR were recently solved from the 90% identical *Fischerella* MV11 HetR, which were determined by X-ray crystallography by the Midwest Structural Genomics Consortium (PDB IDs: 3QOD and 3QOE) (31). In addition to DNA-binding activity, HetR has also been suggested to behave as an unusual serine-type protease with autoproteolytic activity through serine residue S48 (14, 54), although separate mutagenesis studies have made these findings unclear (42). Regardless, HetR is turned over rapidly *in vivo*, presumably by a combination of the inhibitory products of *patS* and *hetN* that are involved in regulating and maintaining heterocyst patterns (10, 51, 52).

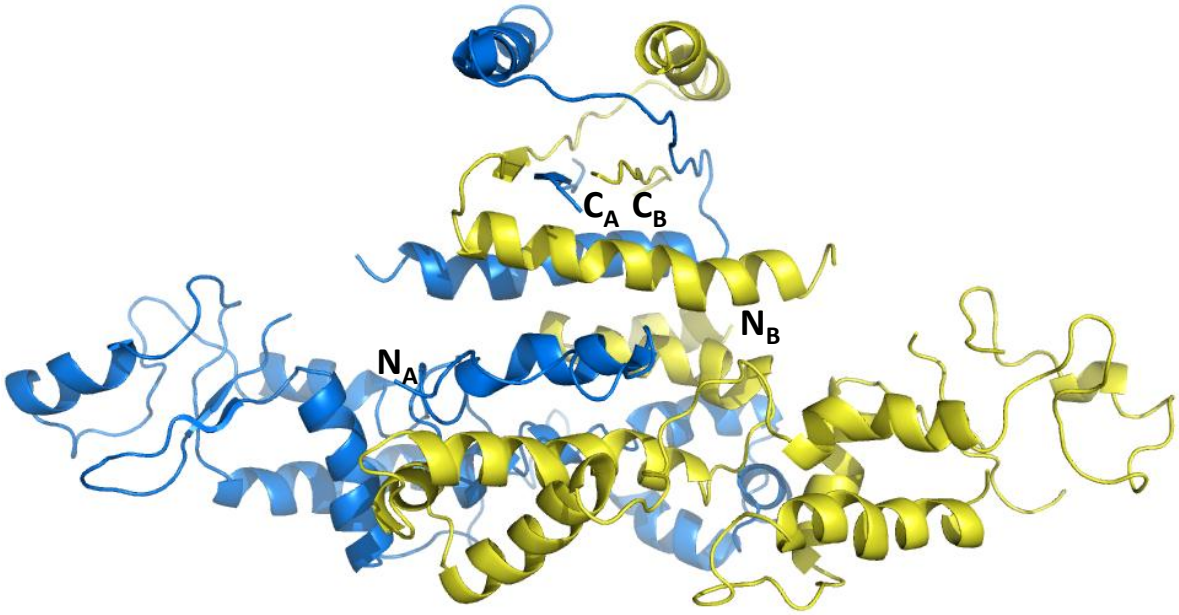

Figure 1.3 X-ray crystal structure of HetR from *Fischerella* MV11 (PDB ID 3QOE) solved as a homodimer at 3.0 Å resolution (31), rendered with PyMOL (12) and shown in a backbone cartoon representation. N- and C-termini for chain A of the dimer (blue) are labeled in the foreground of the structure (N_A and C_A). N-and C-termini for chain B of the dimer (yellow) are represented in the background of the structure (N_B and C_B).

1.2.3 PatS, a heterocyst pattern regulator

The *patS* gene encodes for a 13- or 17-residue (in relation to the two annotated translational start codons) peptide, PatS, in *Anabaena* and is important for controlling heterocyst pattern development (51, 52). Highly expressed in developing cells, proheterocysts, and fully developed heterocysts, functional *patS* must be present in order to maintain the normal 10% pattern of differentiated heterocysts (51, 52). Deletion of *patS* leads to a multiple contiguous heterocyst phenotype of approximately 30% heterocysts, and overexpression leads to complete inhibition of heterocysts even under nitrogen starved conditions (51). It is thought that PatS functions by lateral inhibition of neighboring cells through which a concentration gradient is established along the *Anabaena* filament in order to promote the pattern of one heterocyst cell for every ten vegetative cells. The PatS peptide and the HetN protein, a predicted oxidoreductase from the alcohol dehydrogenase family (5, 10), both contain the amino acid sequence motif RGSGR, which inhibits heterocyst differentiation when supplied ectopically to the culture medium (51), and inhibits HetR DNA-binding activity to its own promoter *in vitro* (29, 42). Mutations to the RGSGR C-terminus of *patS* result in increased heterocyst frequency similar to a *patS* knockout (51). Taken together, the evidence suggests the RGSGR pentapeptide of PatS (PatS-5) contains the critical residues for PatS function, although it has not yet been verified to be the active form of the peptide, since PatS has not been successfully isolated from *Anabaena*. However, Wu *et al*. (48) demonstrated that expression of various *patS* minigene-encoding peptide sequences, ranging in size from PatS-5 (RGSGR), PatS-6 (ERGSGR), PatS-7 (DERGSGR), and PatS-8 (CDERGSGR), could only function cell-autonomously and were unable to laterally inhibit heterocysts, indicating that full length PatS must be processed and exported to neighboring cells for proper pattern regulation. Recent work from Sean Callahan's group have shown that the RGSGR motif of HetN is also required for proper heterocyst patterning (28) and that regulation must be carried out by the formation of PatS and HetN inhibitory gradients along the *Anabaena* filament. The inhibitor gradients serve to

maintain an inverse gradient of active HetR protein by targeting HetR for destruction by post-translational degradation (41).

The results presented in Chapters 2 and 3 of this dissertation demonstrate that various C-terminal PatS peptide fragments bind directly to HetR at a one to one ratio, with varying binding affinities and thermodynamics, and in the absence of DNA. The different PatS sequences analyzed also showed varying inhibitory effects on HetR DNA-binding activity. Also shown is evidence that HetR binds DNA as a homodimer and likely binds one molecule of double stranded DNA substrate per dimer. All of the following results corroborate the hypothesis that full length PatS must be processed to a smaller peptide signal in order to have HetR-dependent regulation of heterocyst patterning. The presented work sets the stage for further *in vivo* and *in vitro* experiments to determine the active forms of the HetR protein and of the PatS signal from *Anabaena*.

1.3 Structural Genomics

1.3.1 The Protein Structure Initiative

In September of the year 2000, the National Institute of General Medical Sciences (NIGMS) at the United States National Institutes of Health (NIH) initiated the Protein Structure Initiative (PSI), which is a massive international campaign of scientific collaborations with the long-term goal of making available the three-dimensional structures of most proteins based on their corresponding DNA sequences. The targets of PSI have primarily been proteins from families without any structural information available to the public. At its inception, the five-year pilot phase of PSI (PSI-1) consisted of four Large-Scale Centers: the Northeast Structural Genomics Consortium (NESG), the Midwest Structural Genomics Consortium (MSGC), the New York SGX Research Center for Structural Genomics (NYSGXRC), and the Joint Center for Structural Genomics (JCSG). Focusing on methods development in addition to protein structure determination using high-throughput X-ray crystallography and solution state nuclear magnetic resonance (NMR) spectroscopy, PSI-1 ran until June 2005 whereupon PSI-2 was

launched (July 2005-June 2010) and has since utilized groups from academia, government, not-for-profit, and private research institutions in order to solve >600 protein structures annually in addition to enhancing preexisting methods and developing new measures to study challenging projects consisting of globular bacterial proteins, integral membrane proteins, eukaryotic proteins, and protein-protein complexes (9). Now since July 2010, the PSI-Biology phase has the ambitious goal of solving the structure of targets with high biological importance and combines the high-throughput system of protein structure determination established in PSI-1 and PSI-2 with studying the corresponding biological function.

1.3.2 The Northeast Structural Genomics Consortium

The NESG, consisting of research groups from Rutgers University, the Robert Wood Johnson Medical School at the University of Medicine and Dentistry of New Jersey, Columbia University, Miami University, the State University of New York at Buffalo, the Hauptman Woodward Research Institute, the University of Toronto, and the University of Georgia, has determined solution NMR and X-ray crystal structures of select proteins ranging from bacteria to humans, with emphasis on targets from large protein families or those with biomedical implications. Development of high throughput parallel technology at the levels of target selection, gene cloning, recombinant protein purification, and sample labeling have also been of significant importance to the NESG in order to increase output and efficiency of the structural genomics pipeline (49).

X-ray crystallography has long been considered the standard method for its uses in drug discovery and drug development research; however, the recent advances in structural genomics have demonstrated the effectiveness of solution NMR for drug research when protein crystals are unobtainable and can report on dynamics information of protein-ligand interactions while in solution (37). NMR has the intrinsic distinction of analyzing molecules in solution instead of relying on the growth and optimization of diffraction-quality crystals, which may prove futile in the case of many macromolecules (36). However not all proteins that can be successfully prepared for NMR with

isotopically-enriched labels (^{2}H, ^{13}C, ^{15}N) provide adequate spectroscopic data sufficient to the successful determination of accurate high-resolution structures. Combining both tools of NMR and X-ray crystallography can provide researchers with a higher success rate of obtaining structural information as well as the advantage of being able to analyze macromolecules in both solution and crystallized states, thus acquiring a more complete understanding of the corresponding structure-function relationship (50).

1.3.3 Target selection

Gene targets are identified and selected by the NESG consortium using predictions from a wide variety of experimental and bioinformatics techniques that are used to predict qualities such as protein domains that exhibit high expression yields, soluble and hydrophobic transmembrane domains, structured and/ or disordered regions, metal-binding sites, and targeting sequences, to ensure greater success rates of high throughput protein structures (1, 49). The targets discussed in this dissertation were all selected during phase PSI-2 of the Protein Structure Initiative and were prepared for NMR studies in a similar fashion. The basic strategy of the NESG for the high throughput production of protein samples to be studied by NMR and determine accurate high-resolution structures will be discussed.

1.3.4 High-throughput cloning, purification, and isotopic labeling of NESG protein targets

The genes for selected protein sequences are routinely cloned from cDNA for eukaryotic targets or from one of many prokaryotic DNA templates in the form of synthetic genome libraries amplified using a Whole Genome Amplification by Multiple Displacement Amplification method for hundreds of bacterial targets (2, 49). Amplified genes are cloned into various expression vectors, for example the pET system of vectors (Novagen), often with C-terminal His_6 affinity tags containing the non-native sequence LEHHHHHH used for later purification. Such vectors are transformed into *E. coli* cell lines, for example: BL21(DE3), which are regularly used for over-expression of proteins

in minimal MJ9 media supplemented with isotopically-labeled $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose, and supplying the lactose analog isopropyl- β-*D*-thio-galactoside (IPTG) to induce expression under control of the *lac* operon. Harvested cell lysates are subjected to affinity column and gel-filtration column chromatography for purification of isotopically-labeled protein samples, which are then concentrated to levels suitable for NMR spectroscopy (~1 mM) and stored in thin-walled Shigemi tubes for NMR data collection. Samples prepared in NMR tubes at Rutgers University, are received at Miami University where they can immediately be placed in the NMR spectrometers for data collection at the High Field Magnetic Resonance Laboratory.

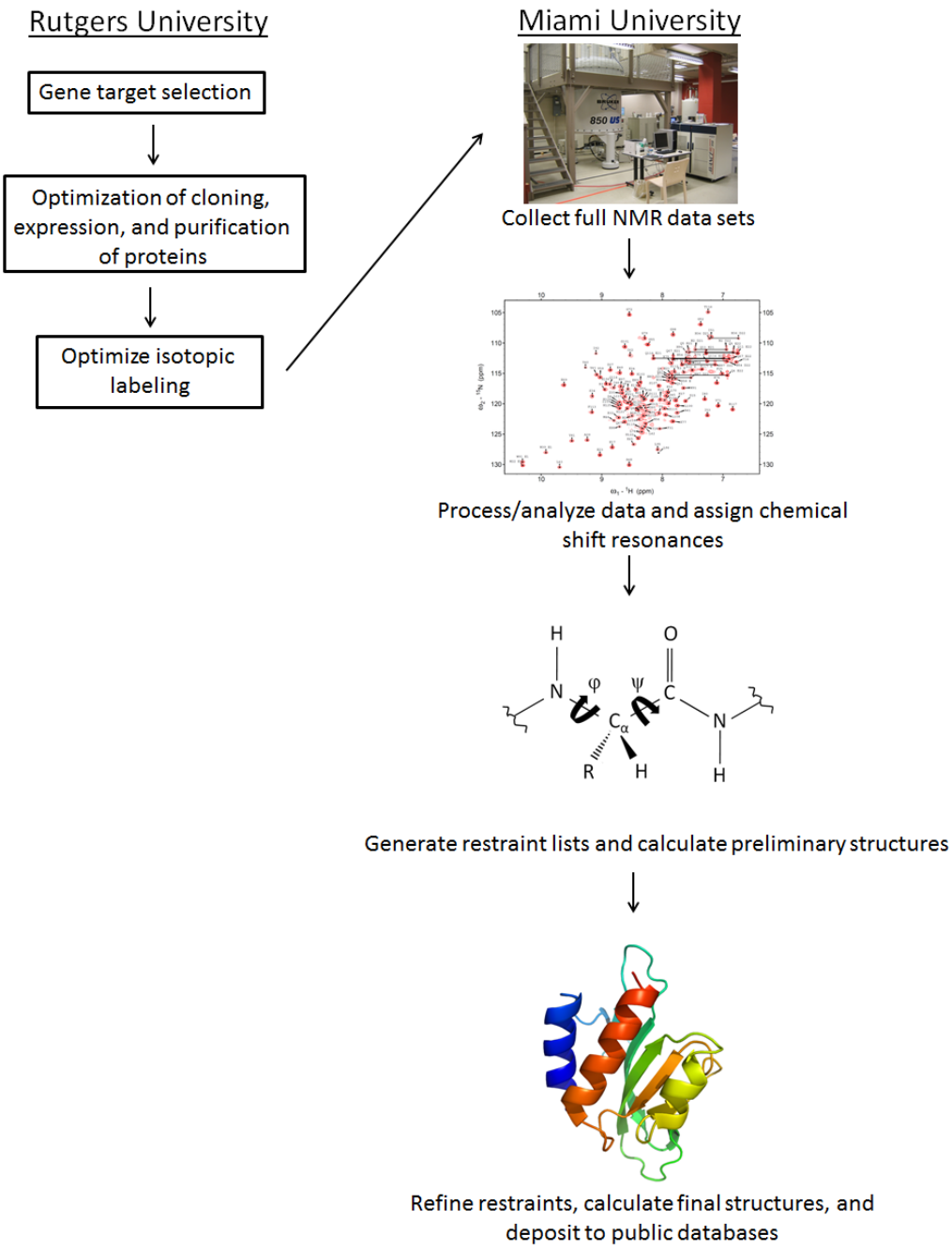

Figure 1.4 Flow chart summarizing the process of high-throughput NMR-based structural genomics at the Miami University Center of Excellence in Structural Biology and Metabonomics for the Northeast Structural Genomics Consortium and funded by the NIGMS (NIH).

1.3.5 Multidimensional Nuclear Magnetic Resonance (NMR) spectroscopy

Isotopically-labeled proteins are analyzed using a combination of 1D, 2D, 3D, and 4D double and triple resonance experiments (typically ^{1}H, ^{15}N, and ^{13}C nuclei) that are carefully optimized for each sample. Chemical shift resonance assignments of the backbone and side chain atoms can then be made for ^{1}H, ^{15}N, and ^{13}C atoms of the protein; the completeness of assignments, however, depends heavily on the quality of the data collected. In one such experiment, the heteronuclear single quantum coherence (HSQC), optimized for detecting proton-nitrogen couplings (^{1}H-^{15}N-HSQC), magnetization is transferred from a ^{1}H nucleus to an attached ^{15}N nucleus by way of the J-coupling through a process referred to as an INEPT or insensitive nucleus enhancement by polarization transfer. Chemical shift evolution is then allowed to take place on the nitrogen, after which the magnetization is transferred back to the hydrogen atoms and the signal detected on the hydrogen (11). This information relates all H-N correlations in the protein, *i.e.*, backbone amide groups and side chain amino groups, and can be used to identify specific amino acid identities when synchronized with other triple resonance experiments, for example those reporting on the coupling of ^{13}C nuclei with the attached ^{15}N nucleus of the H-N pair. Another example of a commonly used experiment is ^{15}N-NOESY, which takes advantage of the Nuclear Overhauser Effect (NOE) where spin magnetization is exchanged between all hydrogen atoms within about 5Å of one another. The hydrogen magnetization undergoes a period of chemical shift labeling and followed by a mixing period during which protons exchange magnetization via the NOE. The magnetization is then transferred from the hydrogen atoms to the directly bonded ^{15}N nuclei via the ^{1}H-^{15}N J coupling, followed by a period of ^{15}N chemical shift evolution, and the magnetization is once again transferred via the INEPT back to the ^{1}H nucleus for detection (11). An important aspect of this technique is that detection of cross peaks is not limited solely to the amino acid directly associated with a given the H-N pair through a chemical bond. Instead cross peaks are observed for all ^{1}H resonances that are close together in space, even if they are distant in primary sequence space, thus providing long-

range structural information. The ^{15}N-NOESY experiment, like ^{1}H-^{15}N HSQC, can also be optimized for ^{13}C-labeled nuclei (^{1}H-^{13}C-HSQC and ^{13}C- NOESY-HSQC). In the case for ^{13}C experiments, the experimental parameters can be further optimized for either aliphatic or aromatic carbon resonances in order to complete assignments for all H-N-C atoms of the protein.

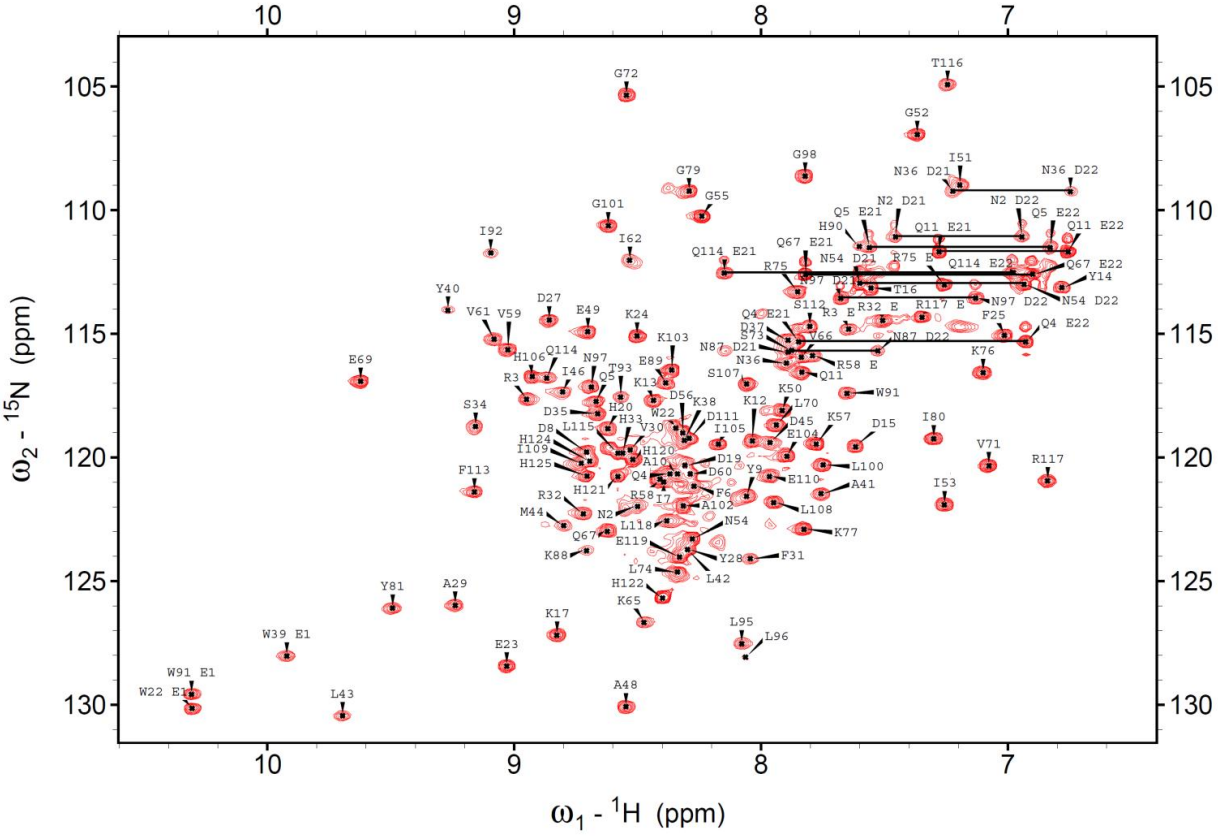

Figure 1.5 Example of a heteronuclear single quantum coherence experiment (^{1}H-^{15}N-HSQC) optimized for H-N correlations for the protein Pspto_3016 from *Pseudomonas syringae*. The assigned resonances for backbone amide and side chain amino H-N pairs are labeled with their respective one-letter amino acid identifiers.

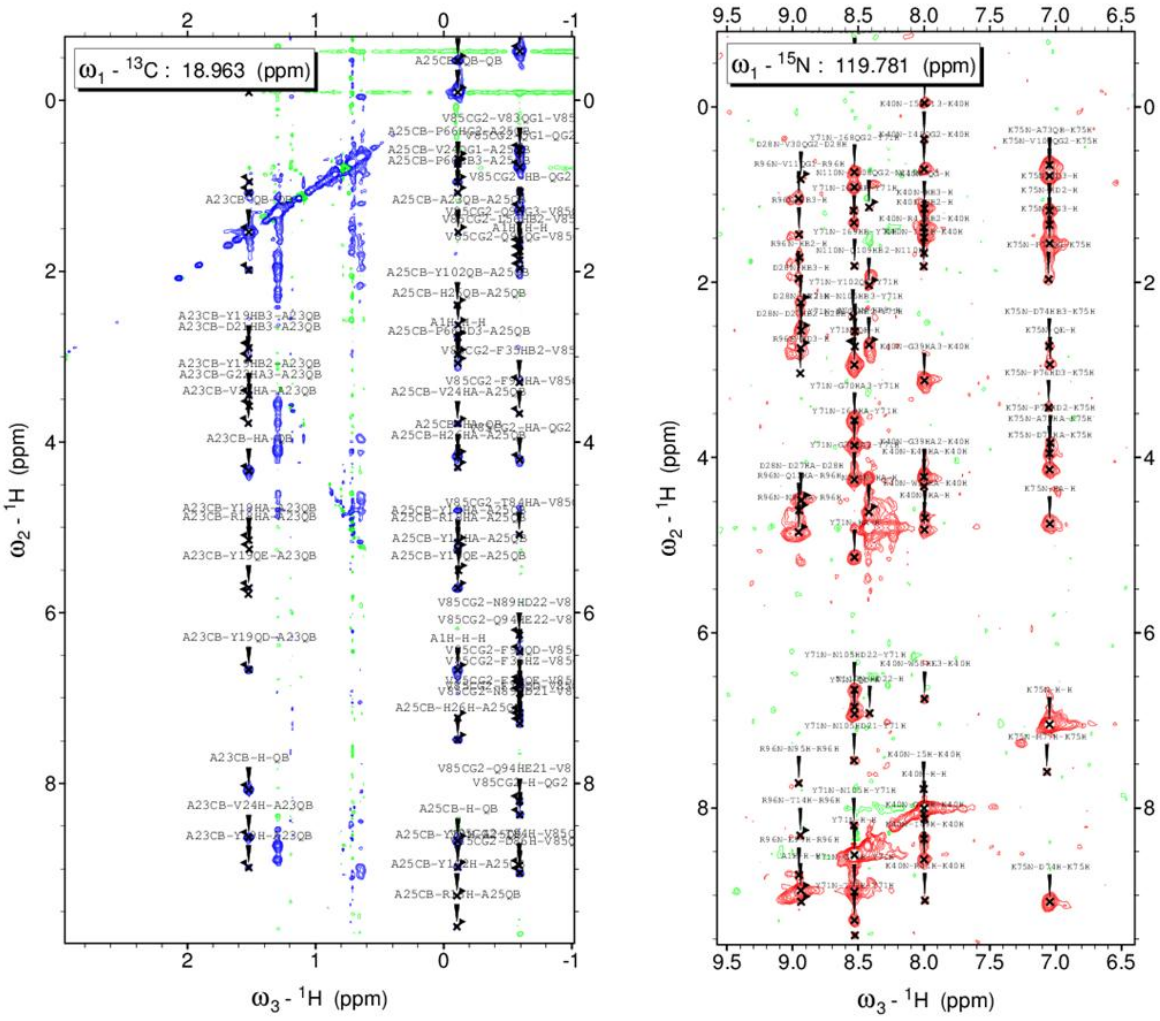

Figure 1.6 Example 2D projections of NOESY spectra (left: aliphatic ^{13}C-NOESY-HSQC; right: ^{15}N-NOESY-HSQC) from the MucBP fragment of the protein LBA1460 from *Lactobacillus acidophilus*. Peaks that are "picked" in the spectrum are indicated with black crosses.

1.3.5 Three-dimensional structure calculations

Resonance assignments for backbone atoms can be used to predict phi and psi dihedral angles of the peptide backbone, which can then be employed as a set of restraints for structure calculations using programs such as CYANA (22, 26). However, it is also necessary to accurately "pick" representative peaks in the NOESY spectra that correspond to nuclei of the protein, while avoiding background peaks due to solvent molecules or experimental noise. The list of "picked" peaks in the NOESY spectra can be sorted by chemical shift frequencies and then supplied to CYANA along with dihedral angle restraints and chemical shift assignments for preliminary three-dimensional structure calculations using automatically generated NOE distance restraints created by CYANA from the input files. This process is further repeated in order to edit and complete chemical shift assignments, validate or reject dihedral angle restraints, and create new restraints files including local hydrogen bond interactions and if applicable, projection angle restraints from residual dipolar coupling data. Multiple rounds of structure calculations are used to continue refinement of restraint lists, chemical shifts, and peak lists, in order to generate an ensemble of low energy three-dimensional models that can be further validated using global quality scores and structure statistics from the PSVS suite (3) to assist in the accuracy of the protein structure.

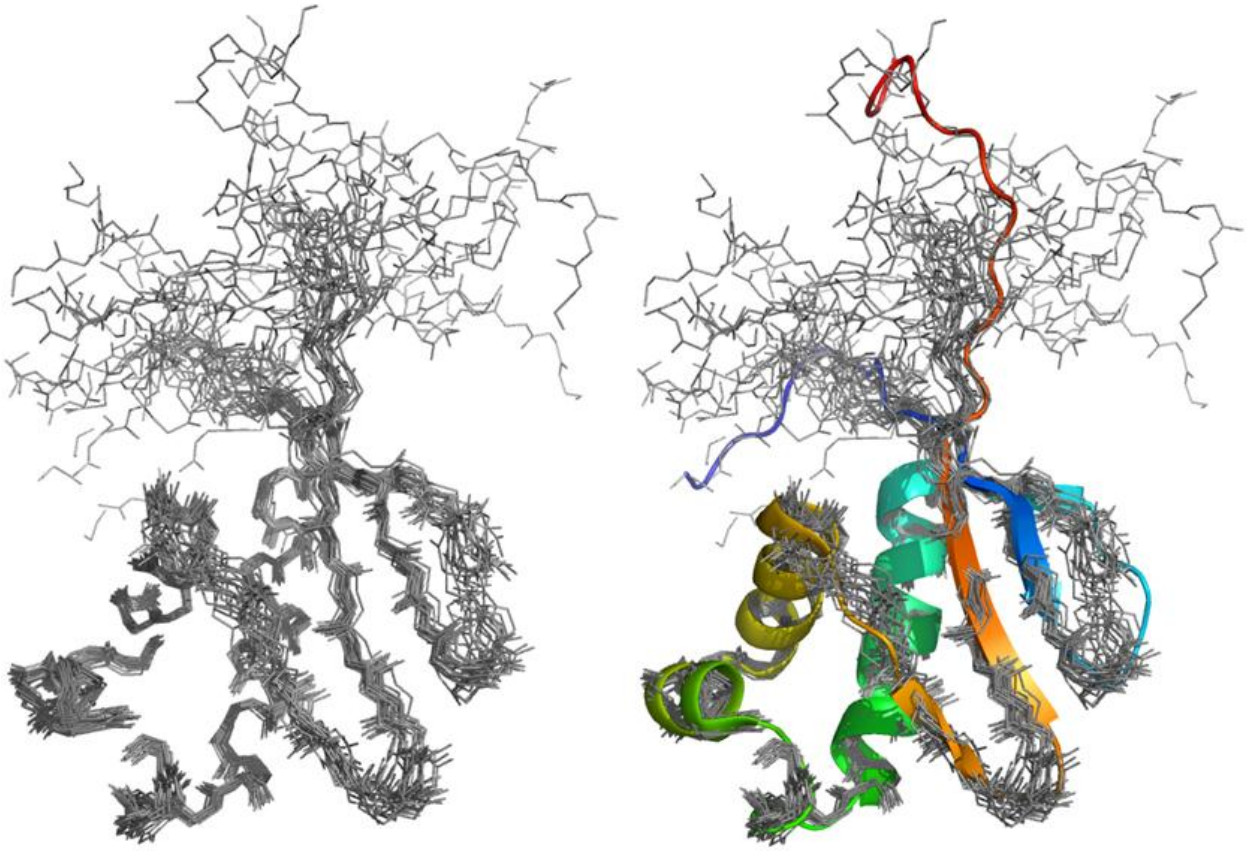

Figure 1.7 Ensemble of 20 structures calculated with the program CYANA for the protein Asl3597 from *Nostoc* sp. PCC 7120. For purposes of clarity, the side chain atoms and protons are hidden. In the right hand ensemble, one model is represented in a cartoon format to assist in identifying secondary structure elements of the protein. The cartoon model is colored from blue to red in the N- to C-terminal directionality and highlights the high ambiguity in atomic position for the disordered termini while the structured core of the protein exhibits significantly more overlap in position of the atomic coordinates for each model.

The results presented in Chapter 5 demonstrate the steps taken to determine high-resolution three-dimensional structures of four bacterial proteins whose structures were previously unknown, using a combination of solution NMR and X-ray crystallography approaches. A combination of different bioinformatics tools was used to identify specific protein functional implications for each protein, two of which were classified as hypothetical proteins at the time of analysis. The structural studies presented here have contributed to the overall goals set out for the PSI, envisioned by the NIGMS (NIH), and will assist in an enhanced biological understanding of the represented systems.

1.4 Dissertation goals and specific aims

Our current knowledge regarding the regulation of heterocyst differentiation and pattern formation have mostly been the result of genetics experiments over the past several decades to identify genes that are important. As the research has continued to grow, an increasing number of genes have been identified, yet the detailed molecular level understanding of heterocyst differentiation is still only little explored. The goals of the research in this dissertation were to characterize the structure-function relationship of the master regulatory protein HetR from *Anabaena*, in order to achieve the necessary atomic level understanding of how HetR initiates heterocyst differentiation in cyanobacteria. In addition, the molecular level of inhibitory regulation by PatS has been characterized to elucidate the means by which heterocyst patterns are regulated in cyanobacteria in order to function as both photosynthetic and nitrogen fixing microorganisms.

The four NESG protein targets discussed in this dissertation were identified from three different bacterial source organisms selected during PSI-2, and their final structures along with raw and processed data used in calculations, were deposited in freely accessible public databases. The goals of the research specifically were to analyze multidimensional solution NMR data collected primarily at Miami University, assign chemical shift resonances for backbone and side chain atoms of each protein and calculate accurate, high resolution three-dimensional structures. In addition, to

collaborate with NMR spectroscopists, crystallographers, and other scientists in order to publish structure reports and include any relevant functional analyses.

Chapter 2 demonstrates the first experimental evidence that HetR is the direct binding target for PatS. Literature reviews have suggested this hypothesis (32, 53) based on the initial report that PatS-5, RGSGR, could inhibit HetR DNA-binding activity to a fragment of its own promoter DNA (29). However this conclusion fails to account for the possibility that PatS-5, based on its charged nature in solution and symmetrically placed arginine residues, could also very likely bind to the DNA in such a way that outcompetes HetR for binding. In this chapter the hypothesis that PatS-5 may bind to HetR, DNA, or both, was tested using a genetics approach in which specific residues in a stretch of the C-terminus of HetR were mutated *in vivo* and the frequency of heterocysts was compared with and without the addition of PatS-5 to the culture medium. Afterward, a site-directed spin labeling approach was used, where the same stretch of residues were individually mutated to cysteine residues in a recombinantly expressed HetR protein from *Escherichia coli* and then spin labeled with the paramagnetic nitroxide radical, MTSL, before probing for perturbations with PatS-5 peptide by both continuous wave and pulsed electron paramagnetic resonance spectroscopy. Lastly, a quantitative approach was tested using isothermal titration calorimetry where PatS-5 was titrated into samples of wild type HetR protein solutions and the thermodynamics of binding were determined. It should also be noted that in this chapter, a method was optimized for expressing and purifying high yields of soluble HetR for subsequent experiments, without the use of denaturation and refolding from insoluble inclusion body fractions that had previously been the standard in the literature (14, 27, 29, 42, 54). In addition, an electrophoretic mobility shift assay method is described for semi-quantitatively determining the binding stoichiometry of HetR for one of its DNA substrates, the double-stranded inverted repeat-containing 29-mer of the *hetP* promoter region. Lastly, indirect evidence is provided for a potential binding site of PatS on the HetR protein based on the qualitative observations from the site-directed spin labeling experiments.

Chapter 3 demonstrates a study of different length fragments from the PatS peptide C-terminus analyzed for their inhibitory effects on HetR DNA-binding activity, and their direct HetR-binding activity. In this study, a spin labeled mutant of HetR, which was described in Chapter 2, was utilized as a probe to detect binding of PatS in order to monitor the interaction by EPR spectroscopy. The fragments of PatS tested consisted of the native *Anabaena* sequences PatS-6 (ERGSGR), PatS-7 (DERGSGR), and PatS-8 (CDERGSGR), and three mutants of the PatS-6 peptide sequence: PatS-6nD (DRGSGR), PatS-6nK (KRGSGR), and PatS-6nG (GRGSGR), which altogether were compared with data collected for the PatS-5 peptide in Chapter 2. The studied provided insight into the physical nature of the PatS:HetR binding interface in the absence of an atomic resolution structure of the complex. Once again, electrophoretic mobility shift assays were used to compare the relative ability of various PatS peptides to inhibit HetR DNA-binding activity for the double-stranded DNA substrate of the 29-mer from the *hetP* promoter region. Continuous wave EPR spectroscopy was used to qualitatively and semi-quantitatively compare the binding interactions of PatS-6 and PatS-8 fragments against the PatS-5 peptide sequence. Lastly, isothermal titration calorimetry was used to directly measure the binding thermodynamics of each peptide to wild-type HetR not only to explore the physical nature of the PatS binding pocket on HetR, but also to learn about the potential requirement for full-length PatS to be processed to a shorter C-terminal fragment in order to function a heterocyst pattern regulator.

Chapter 4 is a collective summary of four independent structural studies of bacterial proteins selected by the Northeast Structural Genomics Consortium and worked on during phases PSI-2 and PSI-Biology of the Protein Structure Initiative. In this chapter, the three-dimensional solution NMR structures and methods used to determine them are described for the following proteins: the putative uncharacterized DNA binding protein Pspto_3016 from *Pseudomonas syringae* (NESG target ID PsR293; PDB ID 2KFP), the hypothetical uncharacterized protein Asl3597 from *Nostoc* sp. PCC 7120 (NESG target ID NsR244; PDB ID 2KRX), the putative uncharacterized protein Asr4154 of the protochlorophyllide reductase superfamily from *Nostoc* sp. PCC 7120 (NESG

target ID NsR143; PDB ID 2L09), and the mucin-binding domain of the protein LBA1460 from *Lactobacillus acidophilus* (NESG target ID LaR80A; PDB ID 2LFI).

1.6 References

1. Acton, T., R. Xiao, S. Anderson, J. Aramini, W. Buchwald, C. Ciccosanti, K. Conover, J. Everett, K. Hamilton, Y. Huang, H. Janjua, G. Kornhaber, J. Lau, D. Lee, G. Liu, M. Maglaqui, L. Ma, L. Mao, D. Patel, P. Rossi, S. Sahdev, R. Shastry, S. GVT., Y. Tang, S. Tong, D. Wang, H. Wang, L. Zhao, and G. Montelione. 2010. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods in Enzymology 493:21-60.
2. Acton, T. B., K. C. Gunsalus, R. Xiao, L. C. Ma, J. Aramini, M. C. Baran, Y. W. Chiang, T. Climent, B. Cooper, N. G. Denissova, S. M. Douglas, J. K. Everett, C. K. Ho, D. Macapagal, P. K. Rajan, R. Shastry, L. Y. Shih, G. V. Swapna, M. Wilson, M. Wu, M. Gerstein, M. Inouye, J. F. Hunt, and G. T. Montelione. 2005. Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. Methods Enzymol 394:210-243.
3. Battacharya, A., R. Tejero, and G. Montelione. 2007. Evaluating protein structures determined by structural genomics consotria. Proteins 66:778-795.
4. Black, T. A., Y. Cai, and C. P. Wolk. 1993. Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in Anabaena. Mol Microbiol 9:77-84.
5. Black, T. A., and C. P. Wolk. 1994. Analysis of a Het- mutation in Anabaena sp. strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing. J Bacteriol 176:2282-2292.
6. Brocks, J. J., G. A. Logan, R. Buick, and R. E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285:1033-1036.
7. Buikema, W. J., and R. Haselkorn. 1991. Characterization of a gene controlling heterocyst differentiation in the cyanobacterium Anabaena 7120. Genes Dev 5:321-330.

8. Buikema, W. J., and R. Haselkorn. 2001. Expression of the Anabaena hetR gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. Proc Natl Acad Sci U S A 98:2729-2734.
9. Burley, S. K., A. Joachimiak, G. T. Montelione, and I. A. Wilson. 2008. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. Structure 16:5-11.
10. Callahan, S. M., and W. J. Buikema. 2001. The role of HetN in maintenance of the heterocyst pattern in Anabaena sp. PCC 7120. Mol Microbiol 40:941-950.
11. Cavanagh, J. 2007. Protein NMR spectroscopy : principles and practice, 2nd ed. Academic Press, Amsterdam ; Boston.
12. DeLano, W., and L. Schrödinger, posting date. The PyMOL Molecular Graphics System, Version 0.99rc6. [Online.]
13. Dixon, R., and D. Kahn. 2004. Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol 2:621-631.
14. Dong, Y., X. Huang, X. Y. Wu, and J. Zhao. 2000. Identification of the active site of HetR protease and its requirement for heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 182:1575-1579.
15. Ehira, S., and M. Ohmori. 2006. NrrA directly regulates expression of hetR during heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 188:8520-8525.
16. Ehira, S., and M. Ohmori. 2011. NrrA, a nitrogen-regulated response regulator protein, controls glycogen catabolism in the nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120. J Biol Chem 286:38109-38114.
17. Ehira, S., and M. Ohmori. 2006. NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium Anabaena sp. strain PCC 7120. Mol Microbiol 59:1692-1703.
18. Fay, P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev 56:340-373.

19. Fogg, G. 1944. Growth and heterocyst production in Anabaena cylindrica Lemm. New Phytologist 43:164-175.
20. Fogg, G. 1949. Growth and heterocyst production in Anabaena Cylindrica Lemm. II In relation to carbon and nitrogen metabolism. Annals of Botany 13:241-259.
21. Frías, J. E., E. Flores, and A. Herrero. 1994. Requirement of the regulatory protein NtcA for the expression of nitrogen assimilation and heterocyst development genes in the cyanobacterium Anabaena sp. PCC 7120. Mol Microbiol 14:823-832.
22. Güntert, P. 2004. Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353-378.
23. Haselkorn, R. 1978. Heterocysts. Annual Review of Plant Physiology and Plant Molecular Biology:319-344.
24. Herrero, A., A. M. Muro-Pastor, and E. Flores. 2001. Nitrogen control in cyanobacteria. J Bacteriol 183:411-425.
25. Herrero, A., A. M. Muro-Pastor, A. Valladares, and E. Flores. 2004. Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. FEMS Microbiol Rev 28:469-487.
26. Herrmann, T., P. Güntert, and K. Wüthrich. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209-227.
27. Higa, K. C., and S. M. Callahan. 2010. Ectopic expression of hetP can partially bypass the need for hetR in heterocyst differentiation by Anabaena sp. strain PCC 7120. Mol Microbiol 77:562-574.
28. Higa, K. C., R. Rajagopalan, D. D. Risser, O. S. Rivers, S. K. Tom, P. Videau, and S. M. Callahan. 2012. The RGSGR amino acid motif of the intercellular signalling protein, HetN, is required for patterning of heterocysts in Anabaena sp. strain PCC 7120. Mol Microbiol 83:682-693.

29. Huang, X., Y. Dong, and J. Zhao. 2004. HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. Proc Natl Acad Sci U S A 101:4848-4853.
30. Kaneko, T., Y. Nakamura, C. P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda, and S. Tabata. 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120. DNA Res 8:205-213; 227-253.
31. Kim, Y., G. Joachimiak, Z. Ye, T. A. Binkowski, R. Zhang, P. Gornicki, S. M. Callahan, W. R. Hess, R. Haselkorn, and A. Joachimiak. 2011. Structure of transcription factor HetR required for heterocyst differentiation in cyanobacteria. Proc Natl Acad Sci U S A 108:10109-10114.
32. Kumar, K., R. A. Mella-Herrera, and J. W. Golden. 2010. Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol 2:a000315.
33. Lambein, F., and C. P. Wolk. 1973. Structural studies on the glycolipids from the envelope of the heterocyst of Anabaena cylindrica. Biochemistry 12:791-798.
34. Laurent, S., H. Chen, S. Bédu, F. Ziarelli, L. Peng, and C. C. Zhang. 2005. Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in Anabaena sp. PCC 7120. Proc Natl Acad Sci U S A 102:9907-9912.
35. Li, J. H., S. Laurent, V. Konde, S. Bédu, and C. C. Zhang. 2003. An increase in the level of 2-oxoglutarate promotes heterocyst development in the cyanobacterium Anabaena sp. strain PCC 7120. Microbiology 149:3257-3263.
36. Montelione, G. T., C. Arrowsmith, M. E. Girvin, M. A. Kennedy, J. L. Markley, R. Powers, J. H. Prestegard, and T. Szyperski. 2009. Unique opportunities for NMR methods in structural genomics. J Struct Funct Genomics 10:101-106.

37. Montelione, G. T., and T. Szyperski. 2010. Advances in protein NMR provided by the NIGMS Protein Structure Initiative: impact on drug discovery. Curr Opin Drug Discov Devel 13:335-349.
38. Muro-Pastor, A. M., A. Valladares, E. Flores, and A. Herrero. 2002. Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. Mol Microbiol 44:1377-1385.
39. Murray, M. A., and C. P. Wolk. 1989. Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two-layers of the cell envelope. Archive of Microbiology 151:469-174.
40. Nicolaisen, K., A. Hahn, and E. Schleiff. 2009. The cell wall in heterocyst formation by Anabaena sp. PCC 7120. J Basic Microbiol 49:5-24.
41. Risser, D. D., and S. M. Callahan. 2009. Genetic and cytological evidence that heterocyst patterning is regulated by inhibitor gradients that promote activator decay. Proc Natl Acad Sci U S A 106:19884-19888.
42. Risser, D. D., and S. M. Callahan. 2007. Mutagenesis of hetR reveals amino acids necessary for HetR function in the heterocystous cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 189:2460-2467.
43. Valladares, A., E. Flores, and A. Herrero. 2008. Transcription activation by NtcA and 2-oxoglutarate of three genes involved in heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 190:6126-6133.
44. Vega-Palas, M. A., E. Flores, and A. Herrero. 1992. NtcA, a global nitrogen regulator from the cyanobacterium Synechococcus that belongs to the Crp family of bacterial regulators. Mol Microbiol 6:1853-1859.
45. Wolk, C. P. 1996. Heterocyst formation. Annu Rev Genet 30:59-78.
46. Wolk, C. P., A. Ernst, and J. Elhai. 1994. Heterocyst metabolism and development, vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.
47. Wolk, C. P., and M. P. Quine. 1975. Formation of one-dimensional patterns by stochastic processes and by filamentous blue-green algae. Dev Biol 46:370-382.

48. Wu, X., D. Liu, M. H. Lee, and J. W. Golden. 2004. patS minigenes inhibit heterocyst development of Anabaena sp. strain PCC 7120. J Bacteriol 186:6422-6429.

49. Xiao, R., S. Anderson, J. Aramini, R. Belote, W. A. Buchwald, C. Ciccosanti, K. Conover, J. K. Everett, K. Hamilton, Y. J. Huang, H. Janjua, M. Jiang, G. J. Kornhaber, D. Y. Lee, J. Y. Locke, L. C. Ma, M. Maglaqui, L. Mao, S. Mitra, D. Patel, P. Rossi, S. Sahdev, S. Sharma, R. Shastry, G. V. Swapna, S. N. Tong, D. Wang, H. Wang, L. Zhao, G. T. Montelione, and T. B. Acton. 2010. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. J Struct Biol 172:21-33.

50. Yee, A. A., A. Savchenko, A. Ignachenko, J. Lukin, X. Xu, T. Skarina, E. Evdokimova, C. S. Liu, A. Semesi, V. Guido, A. M. Edwards, and C. H. Arrowsmith. 2005. NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. J Am Chem Soc 127:16512-16517.

51. Yoon, H. S., and J. W. Golden. 1998. Heterocyst pattern formation controlled by a diffusible peptide. Science 282:935-938.

52. Yoon, H. S., and J. W. Golden. 2001. PatS and products of nitrogen fixation control heterocyst pattern. J Bacteriol 183:2605-2613.

53. Zhang, C. C., S. Laurent, S. Sakr, L. Peng, and S. Bédu. 2006. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. Mol Microbiol 59:367-375.

54. Zhou, R., X. Wei, N. Jiang, H. Li, Y. Dong, K. L. Hsi, and J. Zhao. 1998. Evidence that HetR protein is an unusual serine-type protease. Proc Natl Acad Sci U S A 95:4959-4963.

Chapter 2: Evidence for direct binding between HetR from *Anabaena* sp. strain PCC 7120 and PatS-5

Reproduced with permission from:

Erik A. Feldmann,[1] Shuisong Ni,[1] Indra D. Sahu,[1] Clay H. Mishler,[1] Douglas D. Risser,[2] Jodi L. Murakami,[2] Sasa K. Tom,[2] Robert M. McCarrick,[1] Gary A. Lorigan,[1] Blanton S. Tolbert,[1] Sean M. Callahan,[2] and Michael A. Kennedy*,[1]

[1] *Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056, USA*
2 Department of Microbiology, University of Hawaii, Honolulu, Hawaii 96822, USA

* Corresponding author: Department of Chemistry and Biochemistry, Hughes Laboratories, Room 106, Miami University, 701 High Street, Oxford, Ohio 45056. E-mail: kennedm4@muohio.edu. Phone: 513-529-8267. Fax: 513-529-5715.

This paper has been published in *Biochemistry* 2011, **50**(43):9212-24.

Copyright 2011 American Chemical Society

*Author contributions***:** SN, RMM, GAL, BST, SMC, and MAK contributed to data analysis; IDS, CHM, DDR, JLM, and SKT contributed to data collection and/or data analysis; EAF contributed to *in vitro* data collection and analysis of recombinant HetR proteins cloned, expressed, and purified from *E. coli* including: site-directed mutagenesis, size-exclusion chromatography, MTSL spin labeling, CW and DEER EPR spectroscopy, ITC calorimetry, CD spectroscopy, ICP trace metal atomic emission spectroscopy, and electrophoretic mobility shift assays. EAF also wrote the manuscript.

2.1 Abstract

HetR, master regulator of heterocyst differentiation in the filamentous cyanobacterium *Anabaena* sp. strain PCC 7120, stimulates heterocyst differentiation via transcriptional autoregulation and is negatively regulated by PatS and HetN, both of which contain the active pentapeptide RGSGR. However, the direct targets of PatS and HetN remain uncertain. Here, we report experimental evidence for direct binding between HetR and the C-terminal RGSGR pentapeptide, PatS-5. Strains with a *hetR* allele coding for conservative substitutions at residues 250−256 had altered patterns of heterocysts and, in some cases, reduced sensitivity to PatS-5. Cysteine scanning mutagenesis coupled with electron paramagnetic resonance (EPR) spectroscopy showed quenching of spin label motion at HetR amino acid 252 upon titration with PatS-5, suggesting direct binding of PatS-5 to HetR. Gel shift assays indicated that PatS-5 disrupted binding of HetR to a 29 base pair inverted-repeat-containing DNA sequence upstream from *hetP*. Double electron−electron resonance EPR experiments confirmed that HetR existed as a dimer in solution and indicated that PatS-5 bound to HetR without disrupting the dimer form of HetR. Isothermal titration calorimetry experiments corroborated direct binding of PatS-5 to HetR with a K_d of 227 nM and a 1:1 stoichiometry. Taken together, these results indicated that PatS-5 disrupted HetR binding to DNA through a direct HetR/PatS interaction. PatS-5 appeared to either bind in the vicinity of HetR amino acid L252 or, alternately, to bind in a remote site that leads to constrained motion of this amino acid via an allosteric effect or change in tertiary structure.

2.2 Introduction

Cellular differentiation and patterning are fundamental concepts in the field of developmental biology. One of the earliest known examples of cell differentiation is that of ancient filamentous cyanobacteria. Under pressure from nitrogen starvation, many of these cyanobacterial species evolved the capability to fix atmospheric N_2 using specialized terminally-differentiated cells called “heterocysts” (1, 2, 18, 23-25, 29, 36,

49, 55, 70, 74) more than 2 billion years ago (9, 68). Heterocyst differentiation in filamentous cyanobacteria evolved as a means of isolating oxygenic photosynthesis associated with CO_2 fixation in vegetative cells from oxygen-sensitive nitrogenases that carry out nitrogen fixation in heterocysts (21, 28). Soluble nitrogen-containing compounds generated as a result of nitrogen fixation in heterocysts are shared with neighboring vegetative cells to sustain continued growth of the organism (see discussion by Haselkorn (35) regarding intercellular transport in filamentous cyanobacteria). Under conditions of nitrogen starvation, a pattern is established in which approximately every tenth cell along the filament is terminally-differentiated into a heterocyst (26, 71). Heterocysts support nitrogenase-based N_2 fixation by generating a microoxic environment within the cell that involves production of two additional layers external to the outer membrane found in vegetative cells, including a heterocyst-specific glycolipid layer and a heterocyst envelope polysaccharide layer (44, 50, 59). Initiation of heterocyst differentiation, pattern formation, and pattern maintenance are regulated by small signaling molecules and a host of different genes in a process that resembles signaling pathways of higher eukaryotic organisms (18, 74).

The *hetR* gene plays a central role in regulation of heterocyst differentiation (12). The master regulatory protein, HetR, controls heterocyst differentiation through transcriptional autoregulation (5) and responds to two heterocyst differentiation inhibitors, PatS and HetN, both of which contain the active pentapeptide RGSGR (5, 14, 73). Interestingly, *hetR* and *patS* genes are widespread throughout both non-heterocyst-forming, as well as heterocyst-forming, filamentous cyanobacteria, suggesting that their evolutionary role may have emerged before heterocyst differentiation appeared (75). While much has been learned about regulation of heterocyst differentiation in recent years, the molecular level interactions between HetR, PatS, and HetN during regulation of heterocyst differentiation are still not well understood. There is growing experimental evidence that PatS and HetN control pattern formation by establishing concentration gradients along the filament that promote degradation of HetR in an activator-inhibitor

type manner (64); however, the precise mechanism for how HetR interacts with PatS and HetN remains unknown.

HetR is a 299 residue DNA binding transcriptional regulator believed to be active as a homodimer (39). It has been reported that HetR dimer formation involves a disulfide bridge at position C48 and that mutation of C48 to alanine abolished both dimerization and binding activity to its own DNA promoter *in vitro* (39). It has also been reported that HetR has Ser-type autoproteolytic activity (77, 78) mediated through Ser152 (17). However, Risser and Callahan have reported that Cys48 and Ser152 in HetR from *Anabaena* are not required for proper heterocyst differentiation (65). The cause of the contradictory observations has not yet been resolved.

PatS is a short peptide, predicted to be 13 or 17 amino acids, that acts as a negative regulator of heterocyst differentiation (73). It is thought that a shorter, processed form of PatS acts as a diffusible signal molecule (72, 73). The RGSGR carboxyl terminus of PatS (PatS-5) inhibits heterocyst differentiation when added to culture medium (73) and has been shown *in vitro* to inhibit binding of HetR to a DNA sequence upstream from its own promoter (39, 65). It has also been shown that a R223W mutant of HetR was insensitive to *in vivo* overexpression of both PatS and HetN (45). However, the mechanism of PatS-5 disruption of HetR binding to DNA and the molecular level cause in the loss of R223W HetR sensitivity to PatS remains unknown.

A molecular-level understanding of the biochemical mechanism of action of PatS-5 has not been established. While it has been widely assumed that PatS-5 disrupts HetR DNA-binding through a direct interaction between PatS-5 and HetR, based on observations reported in the literature (39, 45, 65), it is alternatively possible that PatS-5 disrupts the HetR-DNA complex through a direct interaction between PatS-5 and the DNA. This possibility is worth considering given that the amino acid sequence of PatS-5, RGSGR, with its symmetrically positioned arginines, is similar to that of the well-established DNA-binding "AT-hook" motif (63), which has the minimal amino acid consensus sequence PRGRP (4) with both symmetrically positioned arginine residues conferring, and being required for, DNA binding. The AT-hook is the fundamental DNA

binding motif of the HMG-I(Y) (a.k.a. HMGA) subfamily of non-histone high mobility group chromatin proteins (30, 31), also known as "architectural transcription factors" (4, 32, 62), with each HMG-I(Y) protein containing three AT-hooks. When bound to DNA, the AT-hook adopts a crescent-shaped structure that binds AT-rich DNA sequences through non-specific electrostatic interactions between the negatively-charged phosphodiester DNA backbone and the positively-charged arginine side chains that insert into the DNA minor groove (42). Each PRGRP AT-hook motif spans 5-6 DNA base pairs (63). Given the similarity between the amino acid sequences of the AT-hook motif and PatS-5 and the observation that a single AT-hook confers DNA binding capability (4), it is plausible that PatS-5 might also interact with DNA via non-specific electrostatic interactions through its symmetrically-positioned pair of arginine residues. The lack of three-dimensional structures of most heterocyst regulatory proteins and any complexes involving PatS-5 has impeded progress towards achieving a complete molecular level understanding of regulation of heterocyst differentiation. However, the recent report of the crystal structure of HetR from *Fischerella* (46) should begin to open the door to a better understanding of this regulatory process at the molecular level.

Here we report several major findings that should advance the understanding of regulation of heterocyst pattern formation in *Anabaena*, including 1) discovery of a region of HetR including amino acids 250-256 necessary for HetR sensitivity to PatS-5; 2) evidence that PatS binds directly to HetR from a combination of cysteine scanning mutagenesis and continuous wave (CW) electron paramagnetic resonance (EPR) spectroscopy; 3) determination that PatS-5 binds directly to HetR without disrupting the HetR homodimer based on double electron electron resonance (DEER) EPR experiments; and, finally, 4) corroboration that PatS-5 binds directly to HetR with a 1:1 stoichiometry and measurement of the dissociation constant for HetR binding to PatS-5 using isothermal titration calorimetry.

2.3 Materials and Methods

2.3.1 Bacterial strains and growth conditions

Growth of *Escherichia coli* and *Anabaena* sp. strain PCC 7120 and its derivatives; concentrations of antibiotics; induction of heterocyst formation; regulation of the *petE* and *nir* promoters; and photomicroscopy were as previously described (65)**.** Plasmids were conjugated from *E. coli* to *Anabaena* sp. strain PCC 7120 and its derivatives as previously described (19).

2.3.2 Plasmid construction for making chromosomal alleles

Strains of *Anabaena* and plasmids used in this study are described in Table 2.1. Plasmids pJM100, pJM101, pJM102, pJM103, pST211, pJM104, pJM105 and pDR219 are suicide vectors used to replace the chromosomal *hetR*-locus with *hetR(R250K), hetR(A251G), hetR(L252V), hetR(E253D), hetR(E254D), hetR(L255V), hetR(D256E) and hetR(E254G),* respectively. Overlap extension PCR was used to generate each of the mutant *hetR* alleles except *hetR(E254D)* and *hetR(E254G)* that used the inner primers listed in Table 2.2 with names corresponding to that of the resulting substitution with the outer primers PhetR-*BamHI*-F and hetR 3´ Seq. The resulting PCR products were cloned into plasmid pDR327 as *NcoI-SpeI* fragments to create plasmids pJM100, pJM101, pJM102, pJM103, pJM104 and pJM105. Plasmids pST211 and pDR219 were generated in a similar fashion, but outer primer hetR-*SacI-SpeI*-R was used in place of hetR 3´ seq for pST211 and previously published inner primers (65) were used to create *hetR(E254G)* in pDR219.

Strain	Relevant characteristics	Source
PCC 7120	Wild type	Pasteur culture collection
UHM103	*hetR*-deletion strain	(1)
UHM163	*hetR(R250K)*	This study
UHM164	*hetR(A251G)*	This study
UHM165	*hetR(L252V)*	This study
UHM166	Plasmid pDRJM103 recombined into *hetR* locus of strain UHM103	This study
UHM167	Plasmid pST211 recombined into *hetR* locus of strain UHM103	This study
UHM168	*hetR(L255V)*	This study
UHM169	*hetR(D256E)*	This study
Plasmid		
pDR325	Suicide plasmid based on pRL277, carrying *hetR*	(2)
pDR327	Suicide plasmid carrying P_{petE}*hetR-gfp*	(2)
pJM100	Suicide vector carrying P_{hetR}-*hetR(R250K)*	This study
pJM101	Suicide vector carrying P_{hetR}-*hetR(A251G)*	This study
pJM102	Suicide vector carrying P_{hetR}-*hetR(L252V)*	This study
pJM103	Suicide vector carrying P_{hetR}-*hetR(E253D)*	This study
pJM104	Suicide vector carrying P_{hetR}-*hetR(L255V)*	This study
pJM105	Suicide vector carrying P_{hetR}-*hetR(D256E)*	This study
pST211	Suicide vector carrying P_{hetR}-*hetR(E254D)*	This study

Table 2.1 Strains and plasmids used in this study. The sources inducated as **(1)** refer to the studies indicated in Borthakur *et al*. (8) and the sources indicated as **(2)** refer to studies from Risser and Callahan (64).

Oligonucleotide	Sequence 5'-3'
hetRR250K-F	caaacgctatgaaggccttagaagaa
hetRR250K-R	cttctaaggccttcatagcgtttggc
hetRA251G-F	cgctatgcgaggattagaagaactcgatgtgccac
hetRA251G-R	gttcttctaatcctcgcatagcgtttggccg
hetRL252V-F	ctatgcgagccgtggaagaactcgatgtgccacc
hetRL252V-R	cgagttccttccacgctcgcatagcgtttgg
hetRE253D-F	cgcagccttagatgaactcgatgtgccaccag
hetRE253ED-R	catcgagttcatctaaggctcgcatagcgtttg
hetR E254D-F	cgagccttagaagatctcgatgtgccaccagag
hetR E254D-R	gcacatcgagatcttcataggctcgcatagcgtttg
hetRL255V-F	ccttagaagaagtggatgtgccacca
hetRL255V-R	ggtggcacatccacttcttctaaggc
hetRD256E-F	cttagaagaactcgaagtgccaccagagcgctg
hetRD256E-R	ctggtggcacttcgagttcttctaaggctcg
PhetR-BamHI-F	atataggatccaacccttatgacaaaggac
hetR 3' Seq	tgctctacaccacattggttgg
hetR-SacI-SpeI-R	tatatagagctcactagtacttttattcactctgggtgc

Table 2.2 Primers used to generate chromosomal mutants in *Anabaena* sp. PCC 7120.

2.3.3 Construction of strains containing mutant alleles

Strains of *Anabaena* with mutant alleles of *hetR* in place of the wild-type *hetR* were created as described previously (64) using the *hetR*-deletion strain UHM103 and plasmids pJM100, pJM101, pJM102, pJM103, pST211, pJM104, pJM105 and pDR219 to generate strains UHM163, UHM164, UHM165, UHM166, UHM167, UHM168, UHM169 and UHM122, respectively. Strains with *hetR(E253D)* and *hetR(E254D)* are single recombinants in which the entire plasmid is in the *hetR* chromosomal locus, whereas the others are the same as PCC 7120 except for the change in *hetR* sequence.

2.3.4 *In vivo* PatS-5 sensitivity assays

Duplicate cultures of *Anabaena* sp. strain PCC 7120 and the *hetR* mutant strains were grown to an approximate optical density of 0.4 at 750 nm in BG-11 medium, which contains nitrate, a fixed form of nitrogen. For one set of cultures, the culture medium was replaced with fresh BG-11, and replaced thereafter every 48 h with BG-11_0, which lacks fixed nitrogen. For the other set of cultures, PatS-5 was included in the medium at a concentration of 1 μM. The percentage of 500 cells that were heterocysts was determined microscopically and recorded after each change of medium. Reported values are the average of three replicates with one standard deviation. Conditions for photomicroscopy were as described previously (65).

2.3.5 Cloning, overexpression and purification of recombinant soluble HetR

The *hetR* gene was PCR amplified from genomic DNA of *Anabaena* sp. PCC 7120 using the forward primer 5´-ATCGATCGCATATGAGTAACGACATCGATCTGATC-3´ and the reverse primer 5´-TGACTCTCGAGCTAATCTTCTTTTCTACCAAACACC-3´, cloned into pET28b (Novagen) at *NdeI* and *XhoI* sites with an N-terminal 6x-His tag including a thrombin cleavage site, then transformed into competent cells of the cloning host *E. coli* DH5α. Information regarding preparation of *hetR* mutants can be found in Table 2.3. All mutants

were generated using the QuickChange II XL site-directed mutagenesis kit (Stratagene) and their DNA sequences confirmed by capillary electrophoresis based sequencing at the Miami University Center for Bioinformatics and Functional Genomics. HetR 250-256C mutants were generated using the C48A mutant plasmid as the DNA template. Correctly mutated 250-256C plasmids were transformed into BL21(DE3) competent cells containing the pGroESL vector to assist in proper protein folding (52). The plasmid construct was isolated using a Wizard Plus Miniprep kit (Promega) and transformed into competent cells of the expression host BL21(DE3) (Novagen). The *hetR*-containing *E. coli* clone was grown at 37 °C with 250 rpm shaking to an OD_{600} of 0.6-0.9 in 1 L of LB-Miller broth supplemented with 30 µg/mL kanamycin. Protein expression was induced by addition of 0.25 mM isopropyl β-*D*-1-thiogalactopyranoside at 18 °C overnight. Cells were harvested and stored at -80 °C for later use. Thawed cells were resuspended in 25 mL of lysis buffer [1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8] followed by four passes through a French Pressure Cell Press (Thermo Fisher). Cell lysates were centrifuged at 24,000xg for 20 min. The supernatant was loaded onto a 10 mL Ni-NTA affinity column (Qiagen) and washed with 50 mL of wash buffer [1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8] followed by a second wash with pre-elution buffer [1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8, 30 mM imidazole]. Soluble His tagged protein was eluted from the Ni-NTA column with elution buffer [1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8, 300 mM imidazole] and concentrated with an Amicon Ultra (Millipore) to a concentration of 15 mg/mL as determined by the Bradford assay (Thermo Scientific). Concentrated protein solutions were further purified and analyzed on a Pharmacia Superdex200 HiLoad size exclusion column equilibrated with buffer containing 1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8, and 300 mM imidazole using a flow rate of 1 mL/min. The mutants co-expressed with the GroESL chaperone were also supplemented with 30 µg/mL chloramphenicol during expression. Circular dichroism spectra of HetR in the absence of phenylmethanesulfonylfluoride protease inhibitor did not change with the buffer and solution conditions used in our experiments following incubation at 37 °C for 24 hours indicating no loss of structure or change in the

secondary structure composition of the protein due to autoproteolytic activity, which has been reported by Zhou *et al.* (77, 78). Calcium ion concentrations in these solutions were on the order of 50 μM as determined by inductively coupled plasma atomic emission spectroscopy.

HetR mutant	Primer sequences	Cloning host	Expression host	Co-expression vector	Expression temperature
C48A	CAGCAACGGCGGCTAAGGCTGCCATTTACATGAC/ GTCATGTAAATGGCAGCCTTAGCCGCCGTTGCTG	DH5 α	Rosetta(DE3)	none	18°C
R250C	GCGGCCAAACGCTATGTGTGCCTTAGAAGAACTCG/ CGAGTTCTTCTAAGGCACACATAGCGTTTGGCCGC	JM109	BL21(DE3)	pGroESL	22°C
A251C	CGGCCAAACGCTATGCGATGCTTAGAAGAACTCGATG/ CATCGAGTTCTTCTAAGCATCGCATAGCGTTTGGCCG	JM109	BL21(DE3)	pGroESL	22°C
L252C	CCAAACGCTATGCGAGCCTGTGAAGAACTCGATGTGC/ GCACATCGAGTTCTTCACAGGCTCGCATAGCGTTTGG	JM109	BL21(DE3)	pGroESL	22°C
E253C	CGCTATGCGAGCCTTATGTGAACTCGATGTGCCACCAG/ CTGGTGGCACATCGAGTTCACATAAGGCTCGCATAGCG	JM109	BL21(DE3)	pGroESL	22°C
E254C	CGCTATGCGAGCCTTAGAATGTCTCGATGTGCCACCAG/ CGCTATGCGAGCCTTAGAATGTCTCGATGTGCCACCAG	JM109	BL21(DE3)	pGroESL	22°C
L255C	GCGAGCCTTAGAAGAATGCGATGTGCCACCAGAGCGC/ GCGCTCTGGTGGCACATCGCATTCTTCTAAGGCTCGC	JM109	BL21(DE3)	pGroESL	22°C
D256C	GCGAGCCTTAGAAGAACTCTGTGTGCCACCAGAGCGC/ GCGCTCTGGTGGCACACAGAGTTCTTCTAAGGCTCGC	JM109	BL21(DE3)	pGroESL	22°C

Table 2.3 Primers, hosts, and expression conditions used to generate recombinant HetR mutants.

2.3.6 DNA binding assays

Electrophoretic mobility shift assays were performed using 1.8% agarose gels. Agarose gel electrophoresis experiments used buffer containing 0.2 µg/mL ethidium bromide (Fisher) run at 80 mAmp for 1 h in Tris-acetate-EDTA (TAE) buffer**.** DNA binding reactions were incubated for 10 min at 22 °C prior to electrophoresis. Images were generated using an Alpha Innotech camera and Alpha Imager software. The individual strands of the 29 base pair inverted repeat upstream DNA fragment were synthesized (250 nmole scale synthesis for each strand) and HPLC purified (for determining HetR binding stoichiometry), or for all other experiments, synthesized at a 25 nmol scale with standard desalting, by Integrated DNA Technologies (Coralville, Iowa). The complementary oligonucleotides, forward 5´-GTAGGCGAGGGGTCTAACCCCTCATTACC-3´ and reverse 5´ GGTAATGAGGGGTTAGACCCCTCGCCTAC-3´, were annealed by suspending equivalent stoichiometric amounts at 200 µM in the same buffer used to prepare HetR solutions, heated to 85 °C, and then allowed to cool slowly to room temperature. PatS-5 (RGSGR) peptide was custom synthesized and purified by A & A Labs LLC (San Diego, CA), and PolyG (GGGGG) peptide was synthesized and purified by Peptide2.0 (Chantilly, VA), both on a 10 mg scale. PatS-5 was suspended in 100% nanopure H_2O to stock 10 mM concentrations. PolyG was suspended in 100% acetonitrile to a stock 10 mM concentration.

2.3.7 Circular dichroism spectroscopy

Circular dichroism spectra were obtained on a Jasco model J-810 spectropolarimeter. Measurements were obtained using 300 µL of approximately 15 µM recombinant HetR collected at 25 °C using a quartz cell of 1 mm path length. All reported circular dichroism spectra were the result of 10 averaged scans from 200 to 250 nm. Jasco Spectra Analysis software was used to generate the plots of molar ellipticity versus wavelength.

2.3.8 Site-directed spin labeling

The nitroxide spin radical (1-oxyl-2,2,5,5-tetramethyl-pyrrolin-3-yl)methyl methanethiosulfonate (MTSL), (Toronto Research Chemicals Inc.) was dissolved in 50% methanol to a stock concentration of 35 mM. HetR 250-256C mutants were spin-labeled using a two-fold molar excess of MTSL at 22 °C in the dark overnight with gentle shaking. Excess label was removed (confirmed using CW EPR) by size exclusion chromatography using an analytical grade Pharmacia Superdex200 10/300 GL column.

2.3.9 Preparation of samples for electron paramagnetic resonance spectroscopy

Spin-labeled HetR proteins were concentrated to 100 µM. For CW EPR experiments, 30 µL of protein solution was drawn into 1.1 mm internal diameter (1.6 mm external diameter) quartz capillaries. The capillary tubes containing the samples were then placed into 3 mm internal diameter quartz EPR tubes and inserted into the instrument microwave cavity. For pulsed EPR DEER experiments, a cryoprotectant was added to the samples (samples were brought to a final concentration of 30% glycerol) and then 8 µL of the cryoprotected protein solutions were drawn into 1.1 mm internal diameter (1.6 mm external diameter) quartz capillaries. The capillary tubes containing the samples were frozen in liquid nitrogen and then inserted into the resonator for data collection. For DEER experiments involving samples containing DNA, the DNA concentration was 100 µM.

2.3.10 Electron paramagnetic resonance spectroscopy

EPR spectra were collected at the Ohio Advanced EPR Laboratory. CW-EPR spectra were collected at X-band on a Bruker EMX CW-EPR spectrometer using an ER041xG microwave bridge and ER4119-HS cavity coupled with a BVT 3000 nitrogen gas temperature controller (temperature stability ± 0.2 K). CW EPR spectra were collected by signal averaging 15 42-s field scans with a center field of 3370 G and sweep width of 100 G, microwave frequency of 9.5 GHz, modulation frequency of 100 kHz,

modulation amplitude of 1 G, microwave power of 1 mW at 298 K. DEER data were collected using a Bruker ELEXSYS E580 spectrometer equipped with a SuperQ-FT pulse Q-band system and EN5107D2 resonator. DEER data were collected at Q-band with a probe frequency of 34.174 GHz and a pump frequency of 34.235 GHz, a probe pulse width of 20/40 ns, a pump pulse width of 48 ns, shot repetition time of 499.8 µs, 100 echoes/point, 2-step phase cycling at 80 K collected out to 2 ns.

2.3.11 EPR spectral simulations

Qualitative analysis of spin-label mobility was obtained from simulations of the CW EPR line shapes to extract best-fit values of the inverse line-width of the central resonance line (ΔH_0^{-1}) and the components of the diffusion tensor (R) required to reproduce the observed averaging of the hyperfine interaction tensor. The line-width of central resonance line was calculated by measuring central peak-to-peak magnetic field from the first derivative spectrum. Simulations were performed in Matlab using a non-linear least square data analysis program developed by Budil *et al.* (11, 66). The three components of electronic Zeeman interaction tensors ($g_{xx,}$ g_{yy}, and g_{zz}) and hyperfine interaction tensors (A_{xx}, A_{yy}, and A_{zz}) were optimized using the spectrum of HetR 256C, which was characteristic of a spin label approaching the rigid limit. This spectrum provided well-defined features to constrain the A and g tensors to account for the polarity of the local environment of the MTSL nitroxide spin label (38, 76). The A and g tensors were held constant for all remaining simulations and it was assumed that remaining differences in spectra were due to differential motion of the MTSL spin label. The three components of the rotational diffusion tensor (R_{xx}, R_{yy} & R_{zz}) were varied during fitting. The best-fit components of the tensor were averaged to report the overall rate of diffusion ($R_{iso} = 1/3(R_{xx}+R_{yy}+R_{zz})$) (58). EPR spectra of HetR 252C in the presence of PatS-5 consisted of a sum of free and Pats-5 bound species. In this case, a two-site fit was used to account for free and bound states. Best-fit diffusion tensors for each species in the mixture were determined using a Brownian diffusion model. The percentage contribution of each motional component to the overall spectrum was obtained for each sample.

DEER data was simulated using DEER Data Analysis 2009 (43). The distance distributions P(r) were obtained by Tikhonov regularization (15) in the distance domain, incorporating the constraint P(r) > 0. The regularization parameter was adjusted to obtain the realistic resolution.

2.3.12 Molecular dynamics simulations

The atomic coordinates for the HetR crystal structure (PDB ID: 3QOE) from *Fischerella* were downloaded from the Protein Data Bank and used to generate the structures of various spin-labeled HetR constructs with the Nano-scale Molecular Dynamics (NAMD) program (60). The C48A and 250-256C cysteine mutations were created using the molecular graphics software VMD (41). The nitroxide spin-probe MTSL was attached using CHARMM force field topology files incorporated into NAMD. The modified protein assembly was solvated into a spherical water environment and further equilibrated and minimized by running NAMD simulations at room temperature using CHARMM force field parameters. The distance distribution for the 252C mutants was predicted with rotamer library modeling of MTSL conformations using Multi-scale modeling of macromolecular systems (MMM version 2010) (61).

2.3.13 Isothermal titration calorimetry

ITC measurements were performed at 25 °C with a VP-ITC titration calorimeter (Microcal, Northampton, MA). Size-exclusion-chromatography purified HetR samples were dialyzed for 18 h at 4 °C against the reference buffer: 1 M NaCl, 10% glycerol, 300 mM imidazole, 10 mM Tris, pH 7.8. HetR was then diluted to a final concentration of 5 µM for experiments. PatS-5 peptide was diluted into the reference buffer to a final concentration of 100 µM. The titrations were performed in triplicate using a total of 36 injections of PatS-5 peptide for each titration as follows: one 4 µL injection followed by 35 8 µL injections. For each titration 0.60 mL of 100 µM PatS-5 peptide was loaded into the injection syringe. Blank titrations of PatS-5 solutions into reference buffer were performed to correct for the heats of dilution of PatS-5, which were found to be

insignificant. Following the blank experiments, a 1.46 mL sample of HetR was degassed and loaded into the sample cell from the 5 μM prepared stock for each titration. The resulting titration curves were deconvoluted and fit using a one binding site model with the ORIGIN for ITC software package (Microcal, Piscataway, NJ).

2.4 Results

2.4.1 Conservative substitutions at HetR residues 250-256 affect heterocyst formation and sensitivity to PatS-5

As part of a mutagenesis study designed to identify residues of HetR required for function, an allele of *hetR* coding for an E254G substitution was found to cause differentiation into heterocysts of essentially all cells containing a multicopy plasmid carrying the mutant gene (65). Introduction of a plasmid bearing the *hetR(E254G)* allele under the control of the native *hetR* promoter by conjugation resulted in no viable transconjugants. To limit expression of the *hetR(E254G)* allele, the wild-type promoter region was replaced with that of the copper-inducible *petE* promoter, and the plasmid was introduced into PCC 7120. Limited growth of a small number of transconjugants on solid BG-11 medium containing ammonia and lacking copper was observed. The resulting colonies lacked the green color of colonies of the wild-type, were composed primarily of heterocysts, and showed little to no growth in liquid culture (data not shown). By comparison, a strain with P*petE* driving transcription of the wild-type allele of *hetR* in place of the *hetR(E254G)* allele under the same conditions differentiated less than 1% heterocysts. Replacement of the wild-type promoter region with a second promoter, that of *nirA*, from which transcription is repressed in ammonia and induced in nitrate or in the absence of fixed nitrogen (27), permitted the growth of filaments on solid and liquid BG-11 medium with ammonia replacing nitrate as the nitrogen source. Apparently, there is tighter on-off control of transcription with the *nir* promoter than with *petE* in our hands. Transfer of filaments to BG-11 with nitrate or lacking a fixed source of nitrogen resulted in the differentiation of greater than 90% of cells into heterocysts.

By comparison, a strain with P*nir* driving transcription of the wild-type allele of *hetR* in place of the *hetR(E254G)* allele under the same conditions differentiated about 30% heterocysts (data not shown). When the native copy of *hetR* in PCC 7120 was replaced with an allele encoding the E254G substitution, about 25% of cells in the resulting strain were heterocysts 48 h after induction. The phenotype of this strain was indistinguishable from that of a strain with a copy of *hetR* encoding the more conservative E254D substitution, which is discussed in more detail below.

Differentiation of nearly all cells of a filament has been observed when both *patS* and *hetN* are inactivated simultaneously or when an allele of *hetR* encoding protein less sensitive to both inhibitors is overexpressed ectopically and the mutant strains are grown in the absence of combined nitrogen (8, 45). To determine if the more conservative E254D substitution also resulted in an overactive allele of HetR and if residues in the region of E254 were involved in the response of HetR to PatS-5, alleles of HetR encoding individual conservative substitutions at residues R250 – D256 were used to substitute *hetR* by allelic replacement in PCC 7120 (see Table 2.1). Filaments of strains with alleles encoding R250K, E253D, E254D, L255V, and D256E substitutions consisted of about 14 to 48% heterocysts, and the presence or absence of fixed nitrogen in the medium had little effect on differentiation by an individual strain. By comparison, about 9% of cells in filaments of PCC 7120 were heterocysts in a medium that lacked fixed nitrogen, and about 1% when fixed nitrogen was present. Conversely, A251G and L252V substitutions prevented or reduced differentiation, respectively (Figure 2.1). Strains that differentiated an increased number of heterocysts were also less sensitive to PatS-5. Addition of PatS-5 to the growth medium prevented differentiation of heterocysts by the wild-type strain, PCC 7120 (73). In contrast, 8 to 25% of cells in filaments of strains with alleles encoding R250K, E253D, E254D, L255V, and D256E substitutions were heterocysts in a medium that contained PatS-5 at a concentration of 1 μM (Figure 2.1). As expected, PCC 7120 lacked heterocysts under the same conditions. Taken together, these results suggested that residues R250, E253, E254, L255, and D256 of HetR were involved in sensitivity to PatS-dependent signals *in vivo*.

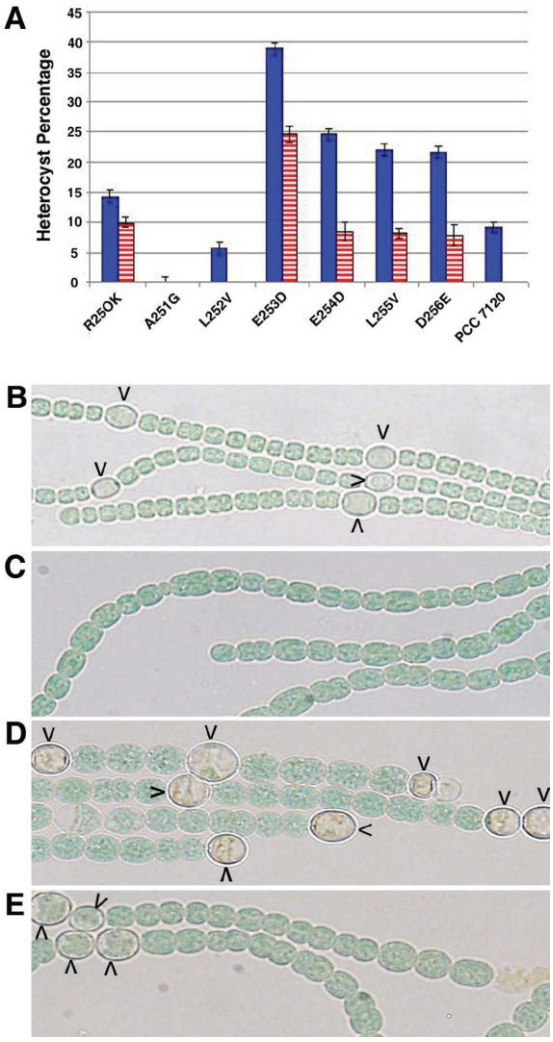

Figure 2.1 Sensitivity to PatS-5 of strains with mutant alleles of *hetR*. (A) Bar graph of the percentage of cells that are heterocysts in the wild-type strain (PCC 7120) and strains with an allele of *hetR* encoding the indicated amino-acid substitution 96 h after removal of combined nitrogen with (red-striped bars) or without (solid blue bars) the addition of PatS-5 to the medium. Values represent the average of 500 cells from three independent cultures. Strain PCC 7120 (B and C) and UHM167, which contained an allele of *hetR* encoding an E254D substitution (D and E), 96 h after removal of combined nitrogen without (B and D) or with (C and E) the addition of PatS-5 to the medium. Carets indicate heterocysts.

2.4.2 Evidence for direct binding of PatS-5 to HetR

The previous experiments pointed to amino acids at the C-terminus as being important for HetR sensitivity to PatS. Guided by these *in vivo* observations, we designed *in vitro* experiments to test the hypothesis that PatS-5 binds to HetR in the vicinity of amino acids 250-256. Initially, we characterized the mobility and secondary structural environment of these amino acids using a combination of site-directed spin labeling and continuous wave (CW) EPR spectroscopy to provide a baseline for PatS-5 binding studies. We then used these techniques to determine how the mobility of these residues were affected by addition of PatS-5 since CW EPR spectra of MTSL spin-labeled proteins can be affected by interactions with other proteins, peptides or small molecules (10, 20, 33, 47, 48). To enable this approach, it was necessary to employ site-specific mutagenesis and nitroxide spin labeling (3), whereby amino acids 250-256 were individually mutated to cysteine and then labeled with the stable nitroxide free radical (1-oxyl-2,2,5,5-tetramethyl-pyrrolin-3-yl)methyl methanethiosulfonate (MTSL). To ensure that MTSL labeling occurred only at one position on the protein, this introduced cysteine needed to be the only cysteine in the protein. Therefore, the HetR used to generate CW EPR spectra had a C48A substitution to remove the naturally occurring cysteine in addition to substituting amino acids 250-256 individually with cysteine. The C48A mutant resembled the unmodified protein in its DNA binding capability and its sensitivity to PatS-5 (see below). The CW EPR spectrum of each MTSL-labeled HetR mutant is shown in Figure 2.2. Control CW spectra of each MTSL spin-labeled mutant were recorded at 100 μM protein, a concentration unlikely to exist in cells but required for measurement of the CW EPR spectra. CW EPR spectra were collected for each mutant after each titration with 100 μM, 200 μM, 500 μM, and 1 mM PatS-5. Simulations enabled determination of g-, A-, and R-tensor parameters (Table 2.4). CW EPR spectra of MTSL labeled at positions R250C and L252C exhibited significant conformational averaging based on their central resonance line widths and isotropic rotational diffusion rates (ΔH_o = 2.03 and 3.45 G, and R_{iso} = 5.37×10^7 and 63.39×10^6 sec^{-1}, respectively) indicating these residues likely occurred either on surface exposed helices or surface

exposed loops (20, 53, 54). MTSL at the A251C, E253C, and D256C positions were less mobile with time-scales of motion too slow to cause motional averaging of the hyperfine interaction tensor (ΔH_o values in the range 5.85-7.70 G and R_{iso} values in the range of $3.47x10^5 – 3.98x10^5$ sec^{-1}) indicating that they likely resided either at helix-helix contacts or in buried locations within the protein (20, 53, 54). The E254C and L255C mutants were difficult to label, exhibited intrinsic protein instability and aggregation, gradually precipitated over time, and produced poor EPR spectra, indicating that mutations at these sites disrupted the local protein structure.

The recently released HetR crystal structure from *Fischerella* (PDB ID: 3QOE, (46)) was used to examine the structural context of the MTSL spin-label in each HetR cysteine mutant. The atomic coordinates for the structure were downloaded from the Protein Data Bank and analyzed using the program VMD (41) in order to generate the C48A mutation, followed by individual R250C-E256C cysteine mutations. Following substitutions, MTSL groups were engineered onto the side chains, attached via the disulfide linkage to the free sulfhydryl of cysteine using NAMD (60) and then simulated to achieve a local energy minimization. The resulting structures are shown in Figure 2.3 and the general location of amino acids 250-256 can be seen in Figure 2.3A. A clear correspondence was evident between the location of each spin label and the magnitude of the conformational mobility based on the inspection and simulation of the CW EPR spectra. For example, the R250C mutant exhibited a highly motionally averaged EPR spectrum (Figure 2.2) that was consistent with the location of residue 250 on the surface of the protein (Figure 2.3B). The A251C mutant displayed an EPR spectrum devoid of motional averaging and the spin label appeared buried in the model (Figure 2.3C). The L252C mutant exhibited significant motional averaging, consistent with the location of the spin label on the surface of the protein (Figure 2.3D). Likewise, the EPR spectra of the E253C and D256C mutants showed no signs of motional averaging consistent with the buried nature of the spin label in the models (Figures 2.3E and 2.3H). Models of E254C and L255C mutants are included in Figure 2.3F and 2.3G for the sake of completeness.

The CW EPR spectrum of HetR spin-labeled at L252C changed dramatically after addition of PatS-5 (Figure 2.4), going from a distinctly motionally averaged spectrum in the absence of PatS-5 to a spectrum that showed no evidence of motional averaging with a ten-fold molar excess of PatS-5, indicating substantial quenching of the conformational dynamics of the MTSL nitroxide electron radical upon addition of PatS-5. These observations provided clear evidence for direct binding between HetR and PatS-5. At PatS-5 to HetR-L252C ratios < 10:1, a mixture of free and bound states in slow exchange was evident and the EPR spectra could be explained as a sum of spectra for free HetR mutant and HetR mutant bound to PatS-5. Simulations of the CW EPR spectra of the MTSL-labeled HetR-L252C mutant enabled determination of the change in the overall diffusion constant for spin-label motion after binding to PatS-5, indicating almost an order of magnitude reduction in the diffusion rate in the bound state ($R_{iso} = 4.17x10^6$ sec^{-1} in the free state compared to an upper limit of $R_{iso} = 8.32x10^5$ sec^{-1} in the bound state; simulation parameters are in Table 2.4). Quenching of the spin-label motion indicated either a proximal interaction with PatS-5 binding nearby, but not necessarily at, amino acid 252, or a distal interaction in which PatS-5 could bind away from the spin label, but the spin label motion could be quenched due to a tertiary or quaternary structural change that caused restriction of the spin label motion at amino acid 252.

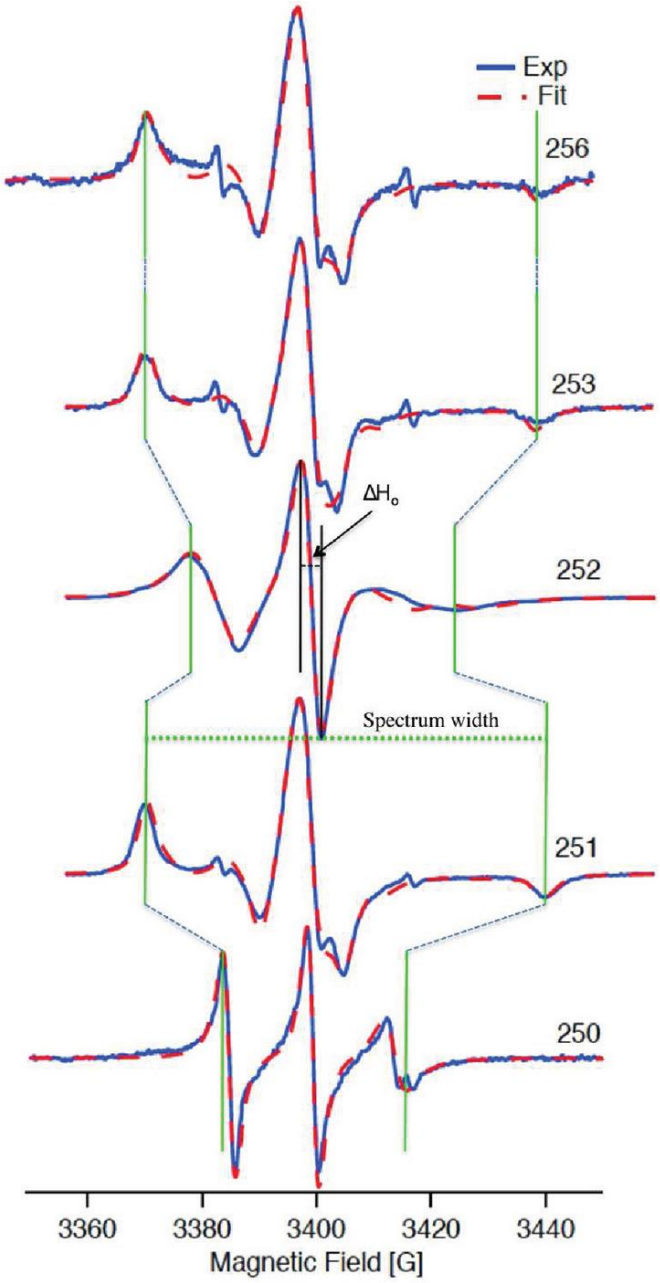

Figure 2.2 Room temperature X-band CW EPR spectra showing variations in spin label mobility for HetR mutants in the range 250-256. The amino acid position of the MTSL label is indicated above each spectrum. The solid blue line indicates experimental data and the red dashed line is the best-fit simulation of the data. Simulation parameters are included in Table 2.4.

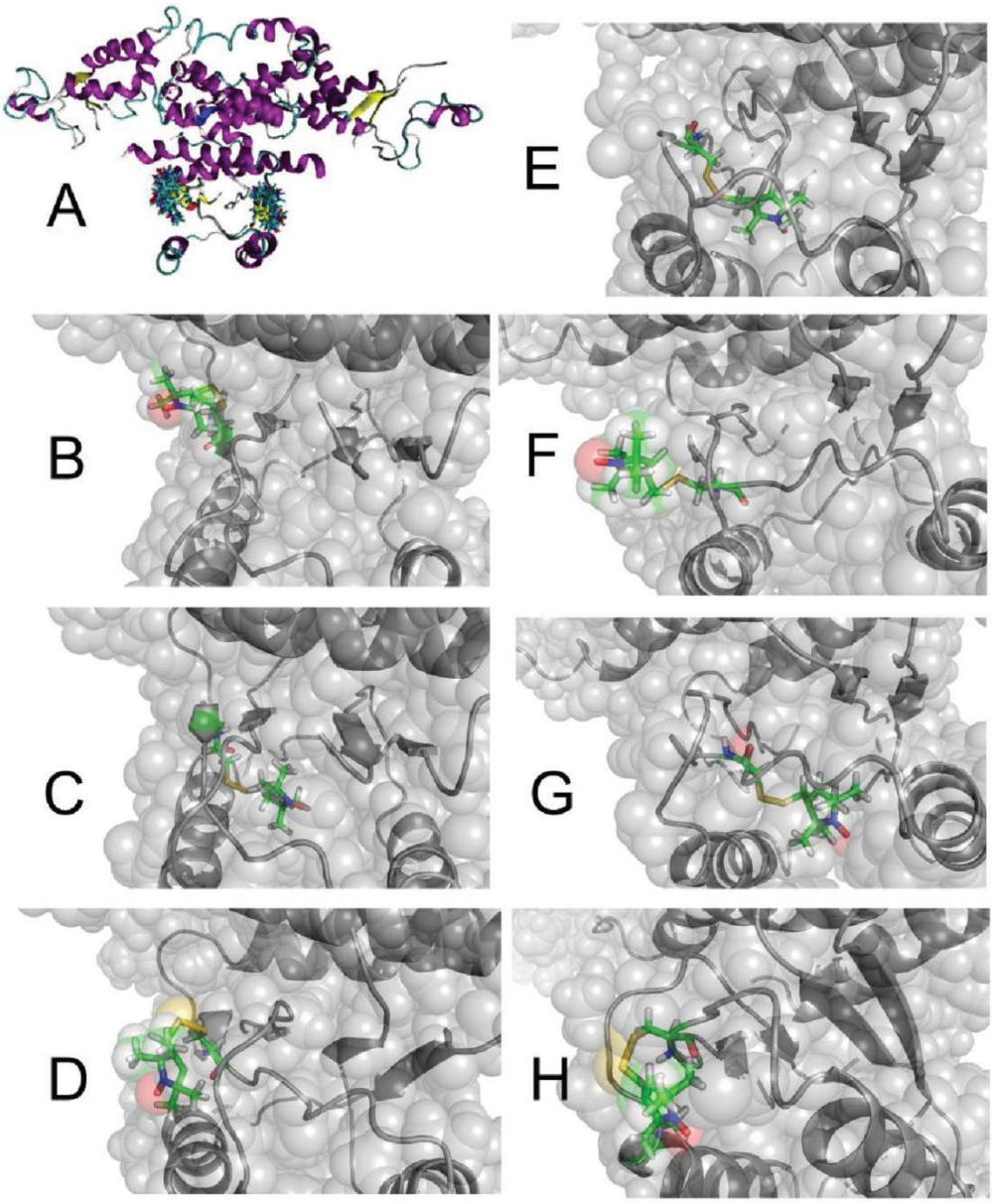

Figure 2.3 A) Crystal structure of the HetR homodimer ((PDB ID: 3QOE, (43)) from *Fischerella* with the general location of residues 250-256 depicted by stick representation which were rendered using Pymol molecular visualization software (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). Space filling representation of the MTSL group in HetR mutant B) R250C, C) A251C, D) L252C, E) E253C, F) E254C, G) L255C and H) D256C.

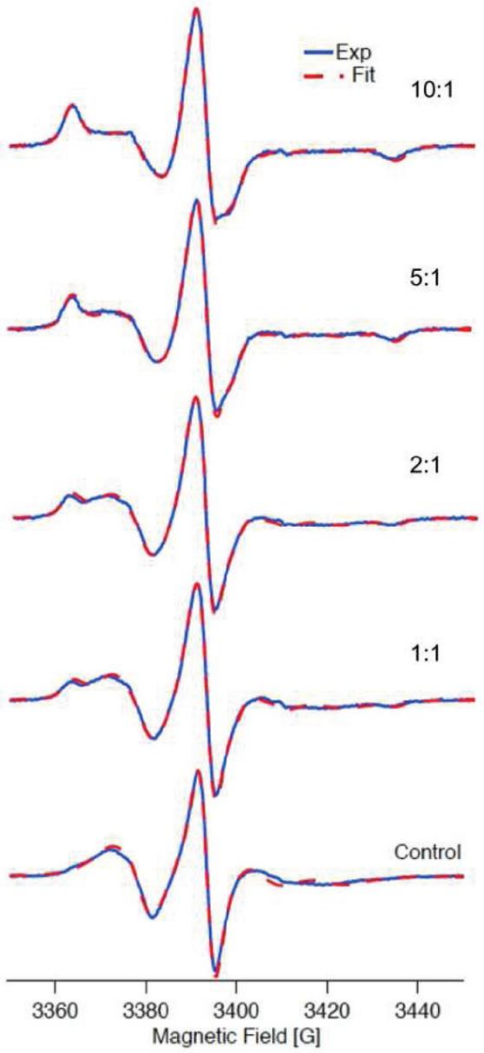

Figure 2.4 CW EPR spectra indicating direct binding between HetR and PatS-5. The MTSL-labeled L252C HetR mutant was titrated with PatS-5 and monitored by CW X-band EPR spectroscopy at room temperature. The control spectrum is shown at the bottom. For the remaining spectra, the ratio of HetR-L252C to PatS-5 (indicated above each spectrum) increases from bottom to top in the stack. The solid blue line indicates experimental data and the red dashed line is the best-fit simulation of the data. Parameters for the EPR spectral simulations are included in Table 2.4.

2.4.3 Evidence for PatS-5 binding to the dimer form of HetR

Pulsed EPR double electron electron resonance (DEER) (56, 57) can be used to measure long-range distances (from 20-70 Å) between spin labels in large proteins (6, 7). Here, DEER EPR experiments were used 1) to measure the intermolecular distance across the HetR homodimer, 2) to determine if PatS binding caused a change in the quaternary structure of the HetR dimer, and 3) to determine if PatS-5 binding disrupted the HetR dimer. DEER data were initially collected on the MTSL-labeled L252C HetR mutant. The time-domain DEER signal (Figure 2.5A) exhibited an excellent signal to noise ratio and exhibited a strong modulation indicating substantial dipolar coupling between the two MTSL spin labels across the dimer. Fourier-transform of the time domain DEER modulation produced a dipolar spectrum with well-defined features (Figure 2.5B). Simulations of the time-domain DEER data produced a clear peak maximum in the distance distribution indicating that the distance between the MTSL spin labels in each monomer was 2.7 nm (Figure 2.5C). Observation of a DEER signal in the MTSL spin-labeled L252C mutant (which contained the C48A mutation) confirmed that C48 was not necessary for HetR dimer formation, since no DEER signal would be detectable if the HetR mutant existed as a monomer in solution. In the presence of a 10-fold excess of PatS-5, the DEER signal was virtually unchanged compared to the HetR alone sample, demonstrating that PatS-5 binds to HetR without disrupting the HetR dimer. Furthermore, the distance between the spin labels, measured to be 2.6 nm in the presence of bound PatS-5, was the same within the uncertainty of the measurement as the distance measured for HetR in the absence of bound PatS-5, indicating that HetR binding to PatS-5 did not cause a structural rearrangement that resulted in a change in the distance between the two spin-labeled L252C residues in the HetR dimer. Furthermore, a DEER distance of 2.7 nm was measured when HetR L252C was bound to the 29 base pair *hetP* DNA fragment, also indicating that the HetR dimer also did not undergo a structural rearrangement that changed the distance between the two L252C residues in the HetR dimer upon DNA binding. The atomic coordinates of the crystal structure of HetR from *Fischerella* (PDB ID: 3QOE (46)) were used to independently measure the distance

across the HetR dimer between residue 252 in the two chains (Figure 2.6). The crystal structure was modified as discussed in the methods section to create the L252C mutant modified by the MTSL spin label. The resulting modified structure was subjected to molecular dynamics simulation to create an ensemble of structures, which are depicted in Figure 2.6A. Distance measurements between the nitroxide groups of the L252C-MTSL modified mutant revealed an average N-N distance of 26.5 Å and an average O-O distance of 27.9 Å (Figure 2.6B), consistent with our observed DEER distance of 27 Å.

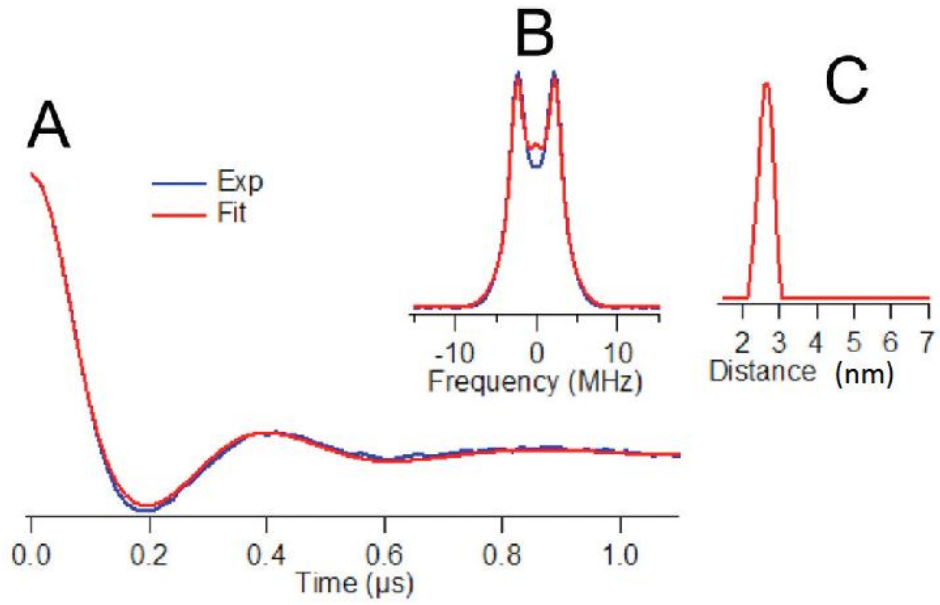

Figure 2.5 DEER data for the L252C HetR mutant. A) the time-domain DEER signal; B) The frequency spectrum resulting from the Fourier transform of the time-domain data; C) the best fit of the distance distribution that explains the DEER data using a Tikhonov regularization fitting procedure.

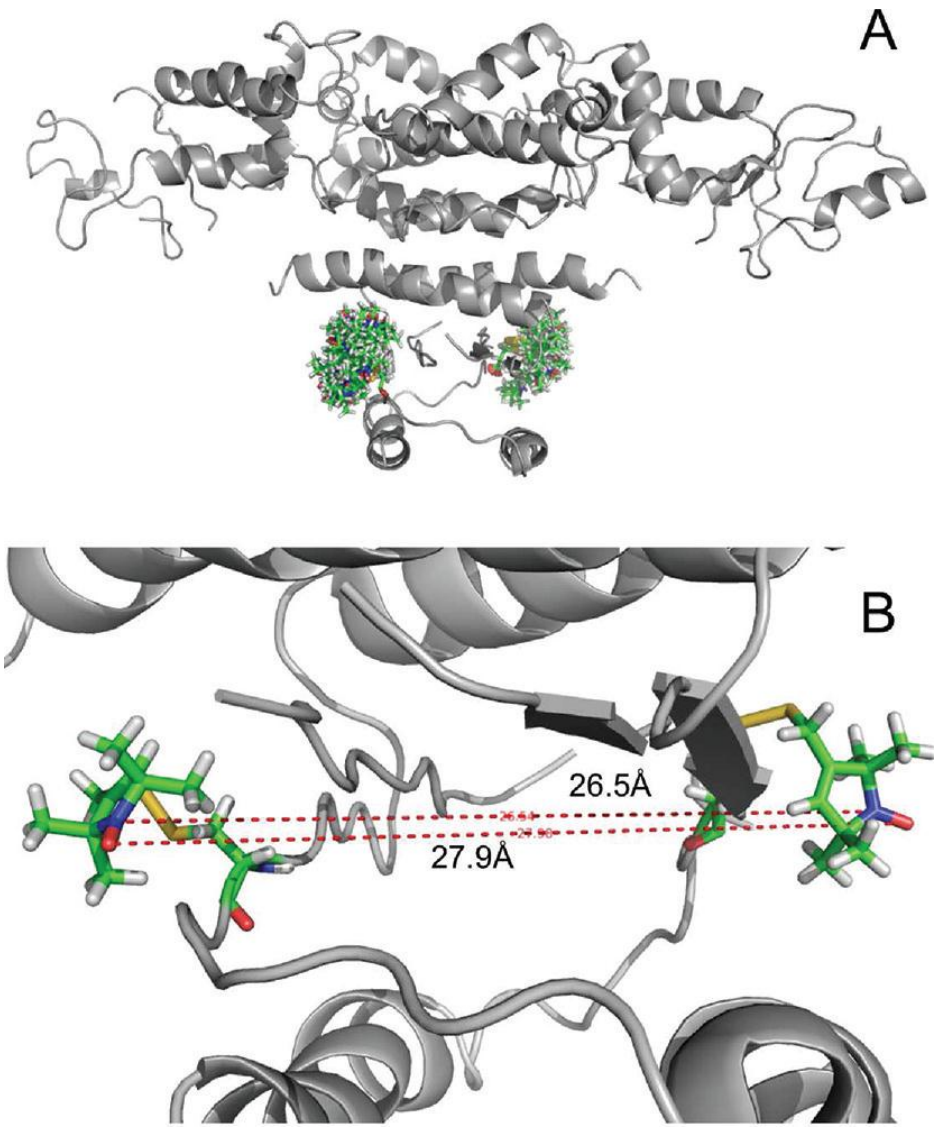

Figure 2.6 A) Superposition of an ensemble of structures of MTSL modified L252C HetR mutant generated by molecular dynamics simulation starting from the crystal structure of the HetR homodimer ((PDB ID: 3QOE, (46)) from *Fischerella* as described in the methods section. B) Representation of the average distance between the nitroxide groups of the L252C-MTSL HetR mutant revealed an average N-N distance of 26.5 Å and an average O-O distance of 27.9 Å.

2.4.4 Characterization of binding of HetR to a 29 bp inverted repeat containing DNA sequence upstream of *hetP*

HetR was recently shown to bind tightly to a 29 bp region upstream of the *hetP* gene containing the inverted repeat sequence 5´-GAGGGGTCTAACCCCTC-3´ (37). Indeed, HetR appears to bind more tightly to this upstream *hetP* DNA sequence than to any previously used DNA substrate reported in the literature, and thus, enabled us to investigate the stoichiometry of HetR binding to DNA. At a HetR to DNA ratio of 3:1, the majority of the DNA was shifted and approximately half of the DNA was shifted at between 2:1 and 3:1 (Figure 2.7). These data are consistent with HetR binding the DNA with a 2:1 stoichiometry, since the extent of completion of the gel shift depends on the dissociation constant, K_d. All these data taken together, along with the fact that most prokaryotic transcription factors bind inverted repeat DNA sequences as homodimers (34, 40, 51), suggested that HetR bound this single inverted-repeat containing upstream *hetP* DNA sequence as a homodimer. Interestingly, at HetR to DNA ratios greater than 4:1, a supershifted species was also present (Figure 2.7), possibly caused by two HetR homodimers binding to a single 29 base pair DNA molecule. It appears that the supershifted species was not caused by non-specific DNA binding since no supershifted species was observed when HetR was mixed with negative control DNA (taken from the human interleuken -2 receptor alpha chain gene promoter PRRII) of the same length and at 100:1 ratio of HetR to DNA (data not shown). The amount of supershifted species depended on the amount of NaCl in solution with the largest ratio of DNA in the supershifted band compared to the shifted band being observed when using 0.5 M NaCl in the solution (data not shown).

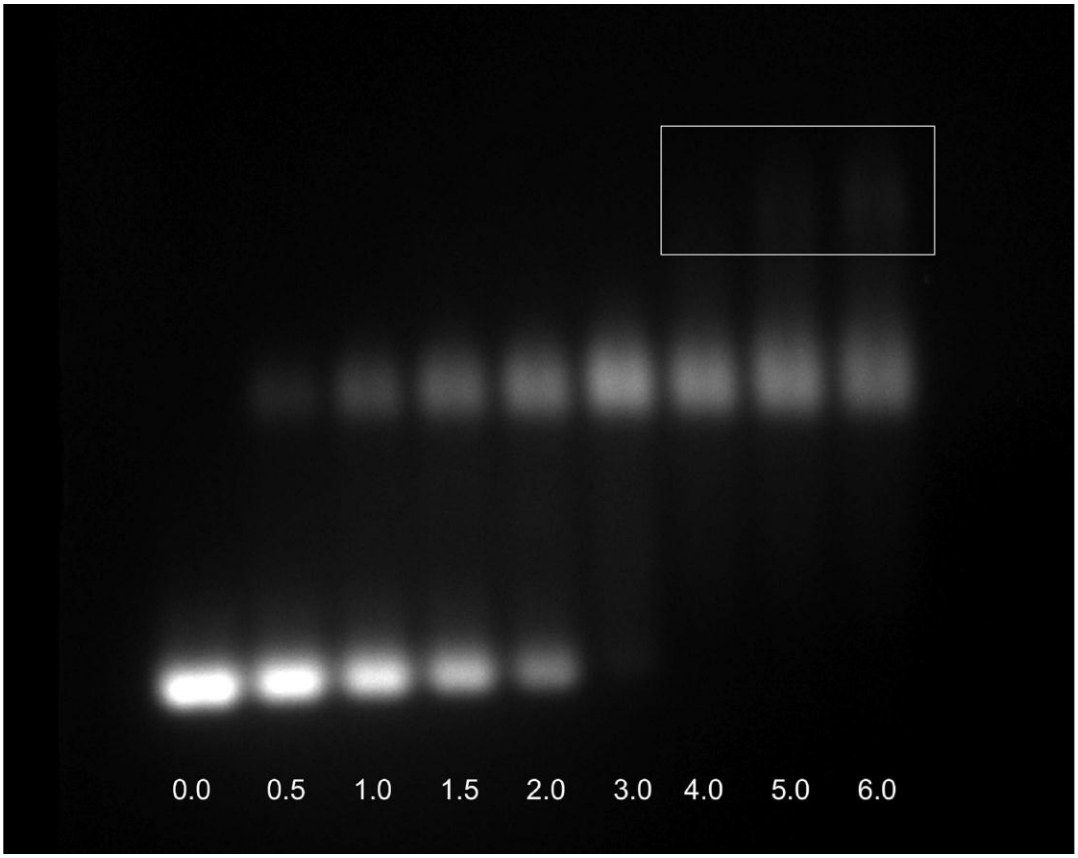

Figure 2.7 EMSA showing binding of HetR to the 29 base pair inverted repeat containing upstream *hetP* DNA fragment. The ratio of HetR to DNA increases from left to right. The resulting completeness of shifted DNA was used to assess the stoichiometry of binding between HetR and DNA as one molecule of HetR dimer per double stranded DNA molecule. The box contains the supershifted bands observed at high HetR to DNA ratios.

2.4.5 Sensitivity of the HetR-DNA complex to PatS-5

The strong interaction between HetR and DNA from the *hetP* promoter region permitted assessment of the PatS-5 to HetR ratios required to disrupt HetR binding to DNA. In gel shift assays we found that PatS-5 completely disrupted the supershifted species at a substoichiometric ratio of 0.5:1 PatS-5 to HetR monomer, and the shifted species was completely disrupted at a 10:1 ratio of PatS-5 to HetR (Figure 2.8). When PatS-5 was titrated into a solution of the MTSL-labeled HetR L252C-DNA complex, the CW EPR spectra were the same as HetR plus PatS-5 alone (Fig 2.9 and Table 2.4). This result was consistent with the gel shift assay showing that the HetR-DNA interaction was disrupted when HetR bound to PatS-5. In other words, in a solution containing HetR, PatS-5 and DNA, only a complex of HetR and PatS-5 would exist and the DNA would be free in solution.

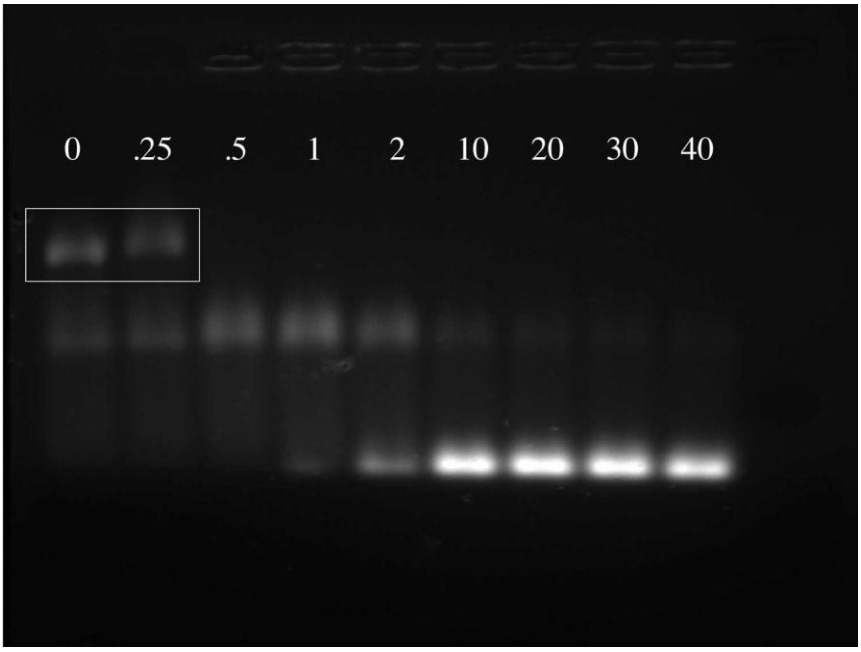

Figure 2.8 Agarose gel shift analysis of PatS-5 disruption of complex formation between HetR and the 29 base pair inverted repeat containing upstream *hetP* DNA fragment. The numbers above each lane indicate the ratio of PatS-5 to HetR monomer. The ratio of HetR to DNA was held constant at 4:1 in each lane. The DNA concentration was 12.5 μM and the HetR concentration was 50 μM. The box indicates the supershifted bands.

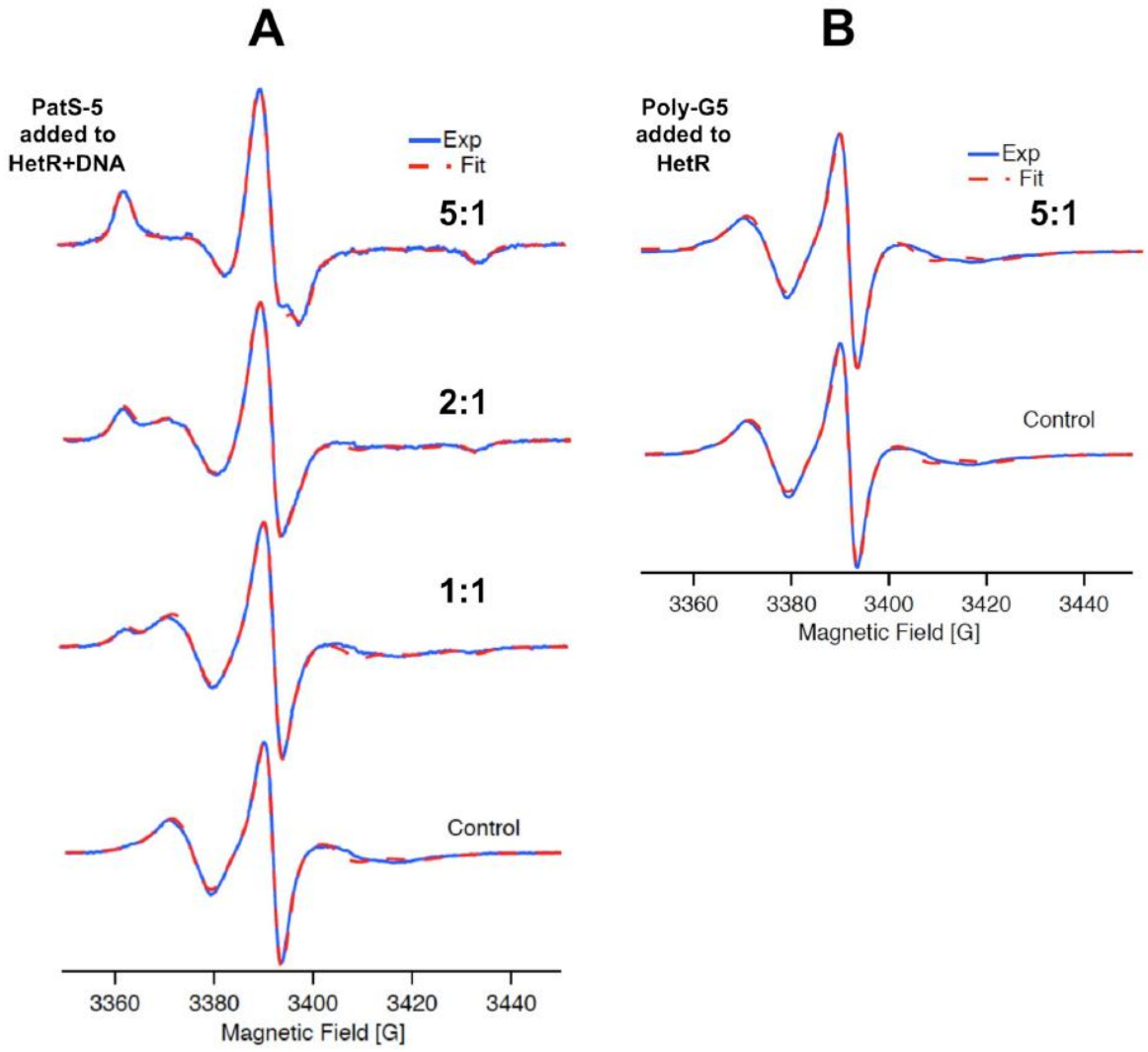

Figure 2.9 Control CW EPR experiments for MTSL-252C-HetR titrations. A) Room temperature X-band CW EPR spectra showing titration of MTSL-252C-labeled HetR with PatS-5 in the presence of DNA. The HetR concentration was 100 µM, the DNA concentration was 100 µM, and the PatS-5 to HetR ratio is indicated above each spectrum. B) Room temperature X-band CW EPR spectra showing MTSL-252C-labeled HetR before and after addition of Poly-G5 pentapeptide. The HetR concentration was 100 µM (control spectrum at bottom), and the Poly-G5 concentration was 500 µM in the top spectrum. The Poly-G5 to HetR ratio is indicated above the top spectrum.

HetR Mutants	Axx	Ayy	Azz	gxx	gyy	gzz	log(Rxx)	log(Ryy)	log(Rzz)	log(Riso)	beta(degree)
250	7.24	5.94	35.55	2.0083	2.0049	2.0023	8.17	6.01	9.01	7.73	22.75
251	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.27	4.60	7.93	5.60	22.48
252	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.73	8.55	4.31	6.53	22.91
253	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.08	4.53	8.12	5.58	22.56
256	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.10	4.60	7.93	5.54	22.46

PatS-5		Axx(G)	Ayy(G)	Azz(G)	gxx	gyy	gzz	log(Rxx)	log(Ryy)	log(Rzz)	log(Riso)	beta(degree)	% site 1	% site 2
100uM_control		7.24	5.94	35.55	2.0083	2.0049	2.0023	7.46	8.34	4.06	6.62	20.78	100.00	0.00
100uM-Pats5	site1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.77	8.53	4.11	6.47	22.21		
	site2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.04	6.20	7.51	5.92	2.10	64.91	35.09
200uM-Pats5	site1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.76	8.53	4.05	6.45	22.68		
	site2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.10	6.22	7.49	5.94	1.76	59.70	40.30
500uM-Pats5	site1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.77	8.50	4.18	6.48	22.09		
	site2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.47	6.16	7.45	6.03	0.72	43.35	56.65
1000uM-Pats5	site1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.77	8.52	4.14	6.47	20.32		
	site2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.35	6.32	7.47	6.05	3.17	26.54	73.46

Table 2.4 Simulation parameters for room temperature CW EPR spectra at X-band.

2.4.6 Similarity of HetR and HetR-C48A

As mentioned above, HetR containing a C48A substitution was used in the CW EPR work that showed binding of PatS-5. It was therefore important to examine whether or not C48A-substituted HetR behaved similarly to wild-type protein. Huang *et al.* (39) previously reported that HetR residue C48 was required for both dimer formation and DNA binding. Here, size exclusion chromatography analysis of HetR-C48A showed that it eluted as a dimer identically to wild-type HetR, indicating that the proposed intermolecular cysteine disulfide bond was not required for dimer formation (Figure 2.10). The DEER EPR data confirmed that the C48A HetR spin-labeled mutant existed as a dimer in solution (discussed above). Comparison of wild-type HetR and HetR-C48A using PAGE (Figure 2.11) showed that the wild-type HetR exhibited a strong dimer band even under denaturing conditions (0.1% SDS in the protein buffer) and in the presence of 10 mM dithiothreitol, whereas the dimer band was almost completely gone in the C48A mutant under the same conditions. Huang *et al.* (39) reported that the dimer band was completely gone for the C48A HetR mutant in their SDS PAGE analysis and concluded that HetR-C48A was unable to form dimers *in vitro*, that the HetR dimer was formed through a disulfide bond, and that Cys48 was required for dimerization. We were, however, able to detect a dimer band for HetR-C48A under milder denaturing conditions using 0.01% SDS-PAGE (Figure 2.11). Taken together with the size exclusion chromatography and DEER data, these results indicated that C48 was not required for dimer formation; however, the dimer still may be stabilized by an interchain disulfide bond. In addition, circular dichroism (CD) spectra of wild-type HetR and the C48A HetR mutant were identical and characteristic of a protein with predominantly α-helical secondary structure as indicated by distinct minima in molar ellipticity at 209 and 222 nm (wild type HetR CD spectrum shown in Figure 2.12).

In addition to having the same physical characteristics as the wild-type protein, C48A substituted HetR behaved similarly to the wild-type HetR in gel shift assays. At the same concentrations as those described above, both shifted and supershifted species were observed; however, the supershifted band tended to be less prominent compared to

that generated with wild-type HetR. The HetR-L252C mutant also shifted DNA in the same manner as the C48A mutant of HetR (data not shown). When PatS-5 was included in the binding reactions, binding of the C48A HetR mutant to the 29 bp *hetP* promoter region was affected in a manner similar to that for the wild type protein, suggesting that the C48A substitution does not affect binding of HetR to DNA or its interaction with PatS-5 peptide.

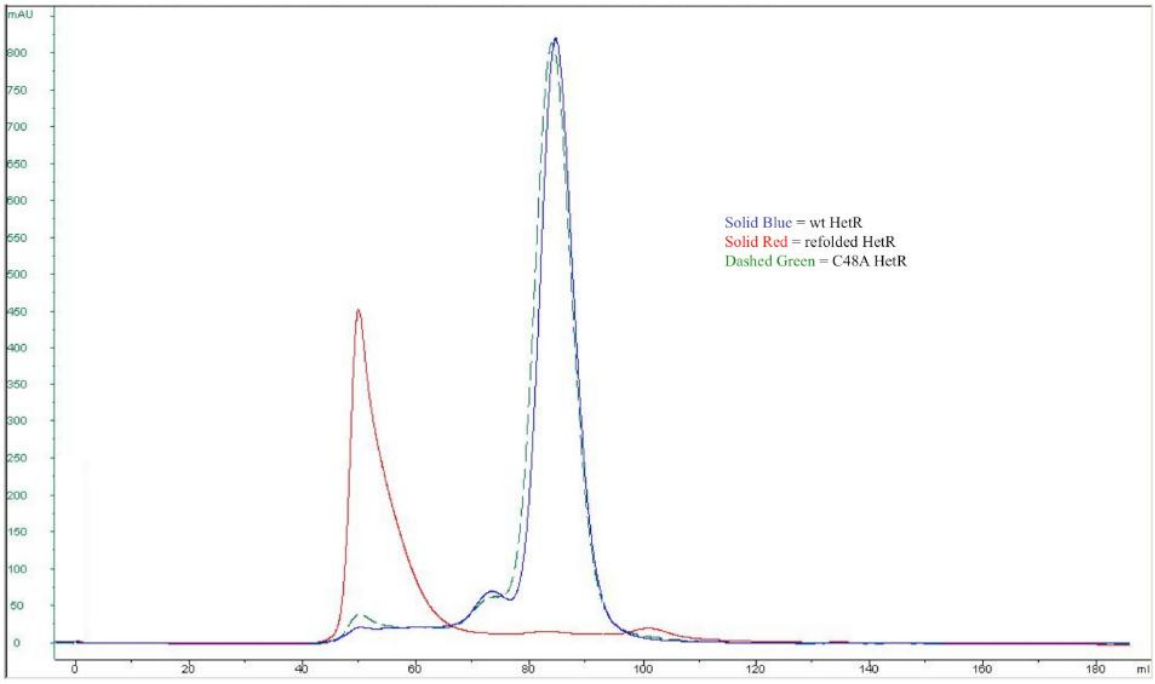

Figure 2.10 Size exclusion chromatogram of HetR protein preparations. Wild type HetR (solid red) elutes off the column at approximately 85 mL, which is identical for the C48A mutant (dashed green). The denatured/refolded HetR protein elutes as a much higher molecular weight aggregate at approximately 50 mL. The Superdex200 column was calibrated with a molecular weight standard kit containing the following proteins: Aprotinin = 6.5 kDa, Ribonuclease A = 13.7 kDa, Carbonic Anhydrase = 29 kDa, Ovalbumin = 43 kDa, Conalbumin = 75 kDa, and Blue Dextran 2000 = 2,000 kDa. A pure sample of BSA (66.4 kDa) was also run as a standard and elutes nearly identical to HetR at 85 mL. The numbers indicated along the y-axis are in the units of milli absorption units (mAU). Units along the x-axis are in mL and refer to the elution volumes.

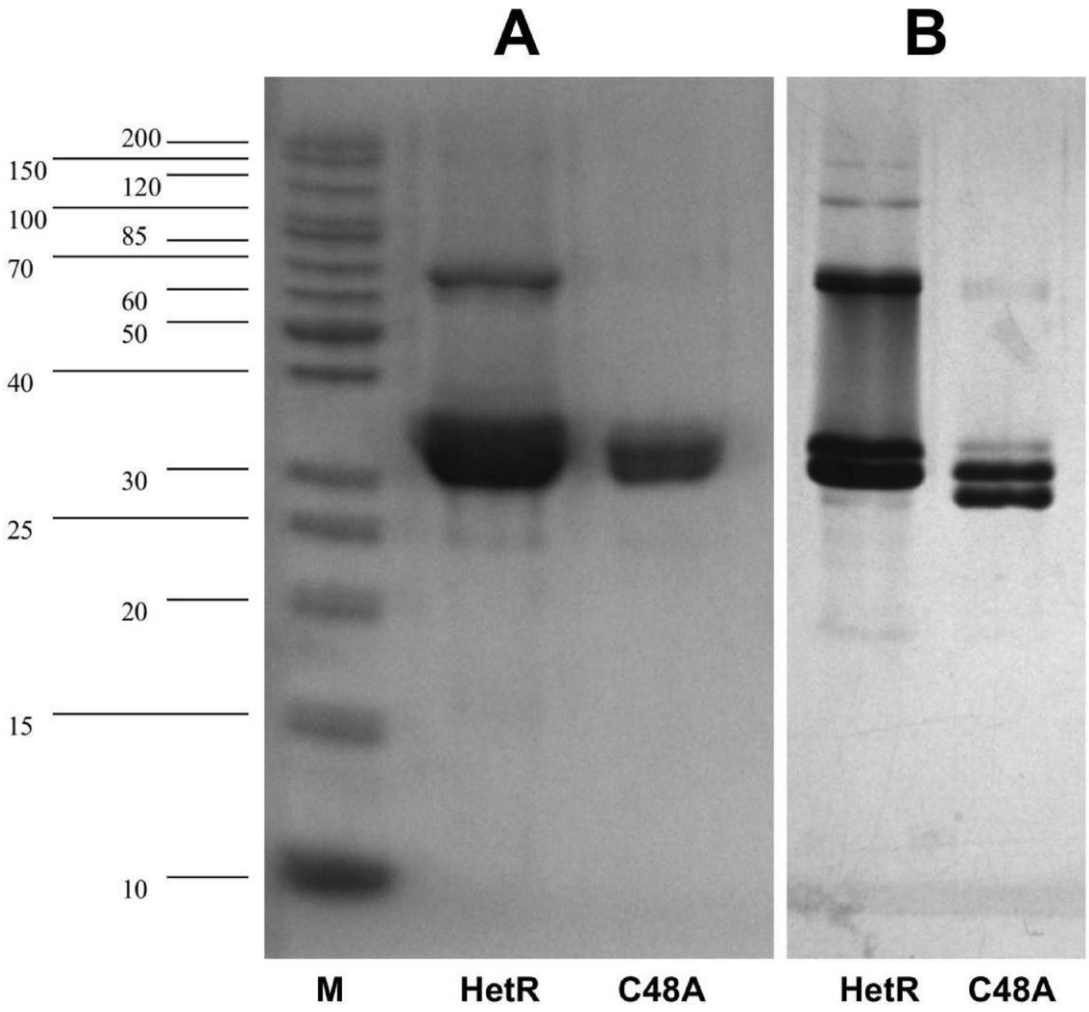

Figure 2.11 Polyacrylamide gel electrophoresis analysis of soluble-expressed purified recombinant HetR and C48A HetR mutant. A) 0.1% SDS and 10 mM dithiothreitol. Molecular weight markers (Fisher Bioreagents EZ-run™ *Rec* Protein Ladder) are indicated in units of kDa to the left of the lane marked M. B) 0.01% SDS and 10 mM dithiothreitol. Lanes are marked "HetR" for wild type HetR and "C48A" for the C48A mutant of HetR.

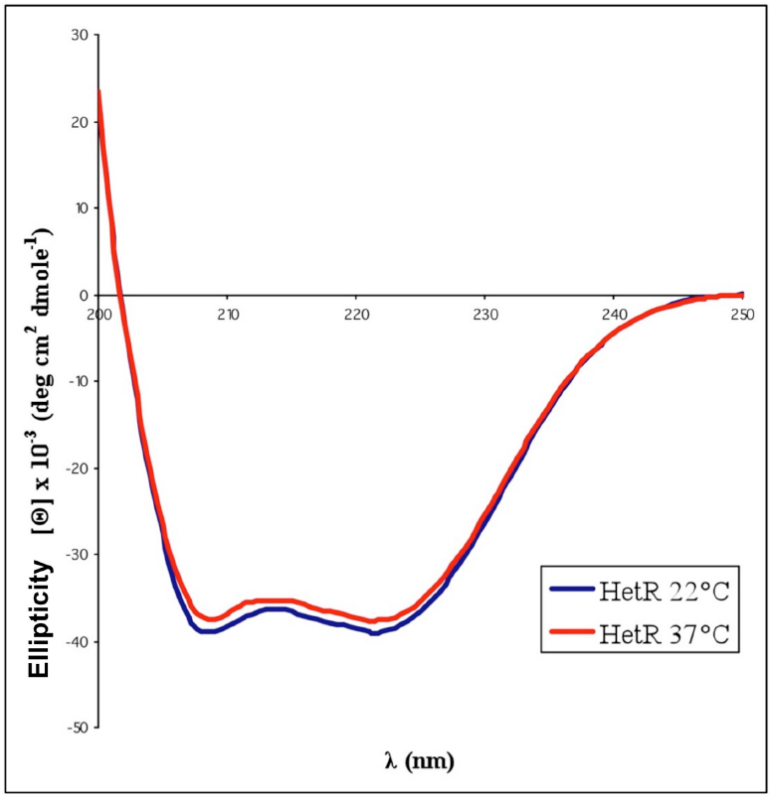

Figure 2.12 Circular dichroism spectra of wild type HetR at room temperature (blue) and after incubation at 37 ºC overnight (red). Based on inductively coupled plasma measurements for metal content in these solutions, the calcium ion concentrations were on the order of 50 µM and the protein concentrations were on the order of 15 µM.

2.4.7 Determination of the stoichiometry, dissociation constant, and thermodynamic parameters for binding of PatS-5 to HetR

Isothermal titration calorimetry (ITC) is a highly sensitive biophysical technique for analyzing thermodynamics of molecular interactions. ITC can be used to study protein-protein or protein-small molecule binding interactions and is capable of detecting heats of binding directly, without use of molecular labels or tags that might interfere with, or influence, binding (22). Here, ITC was used to confirm direct binding of PatS-5 to HetR (Figure 2.13), and the integrated binding heats indicated a single exothermic binding event. A single binding site model was used to fit the binding curves, which allowed determination of the binding stoichiometry *n*, K_a (equal to $1/K_d$), and ΔH. The ΔG was determined from K_a, and ΔS was then calculated using the ΔG *and* ΔH values. The data indicated that one PatS-5 peptide bound to one HetR monomer, *i.e.*, two peptides bind per homodimer. Calculations of the thermodynamic parameters using a model for two non-interacting identical binding sites in a homodimeric protein should be equivalent to those obtained from calculations a single peptide binding-site on a monomer (67). PatS-5 was found to bind HetR with a mean K_d of 227 ± 23 nM, ΔG of -9.063 ± 0.059 kcal/mol, and *n* of 1.04 ± 0.01 sites, indicating a single tight exothermic binding interaction. PatS-5 binding to HetR exhibited a relatively large negative enthalpy (-8.337 ± 0.410 kcal/mol) and a positive entropy (2.44 ± 1.55 cal/(mol·deg)). Inspection Table 2.5 shows that the binding affinity of HetR for PatS-5 was dominated by the change in enthalpy in comparison to the relatively small change (more than 10-fold smaller) in $T\Delta S$. This observation indicated that binding of PatS-5 to HetR was enthalpically-, as opposed to entropically-, driven. It is difficult to elucidate the relative contributions to PatS-5 binding; however, the relatively large negative enthalpy of binding is consistent with the formation of hydrogen bonds or ionic interactions, which are associated with relatively large negative contributions to enthalpy for ligands binding to proteins (13, 16, 69). Collectively, the isothermal titration calorimetry data corroborate the EPR observations that PatS-5 binds directly to HetR in the absence of DNA.

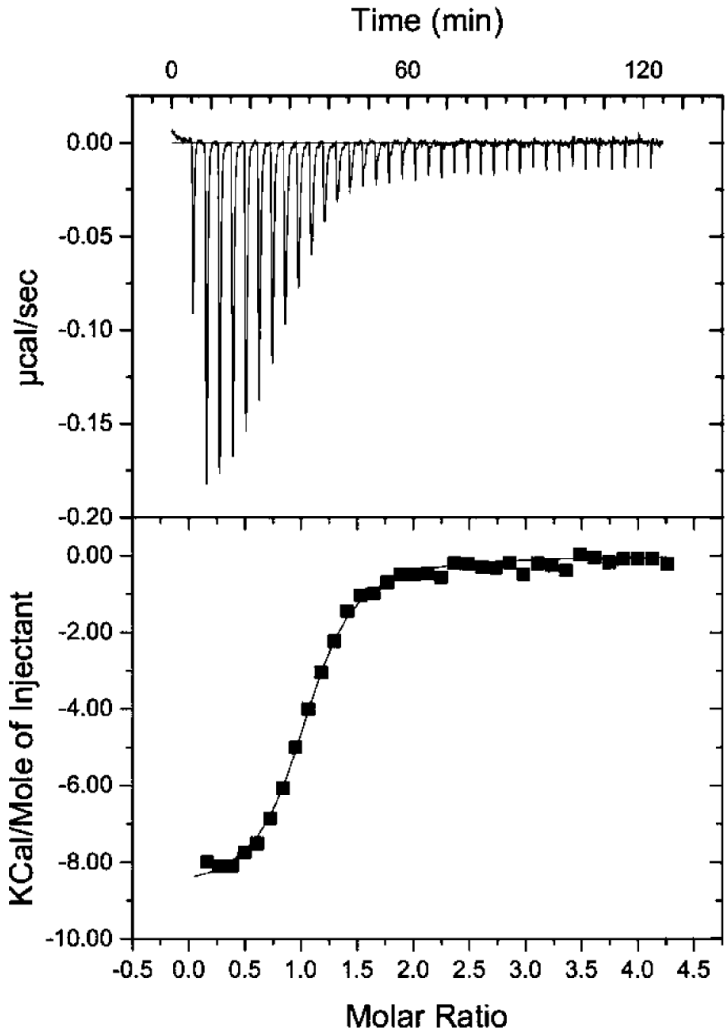

Figure 2.13 Representative ITC isotherm for HetR binding to PatS-5. Experiments were performed in triplicate. Shown are the heat effects as a function of time (µcal/sec) and the cumulative heat effects as a function of the molar ratio of PatS-5 peptide to HetR (kcal/mol). Solid line represents the fit to the experimental data.

peptide	n (no. of sites)	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)
PatS-5	1.04 ± 0.01	227 ± 23	-9.063 ± 0.059	-8.337 ± 0.410	0.73 ± 0.46

Table 2.5 Summary of quantities associated with HetR binding to PatS-5 peptide derived from isothermal titration calorimetry experiments

2.5 Discussion

Since the original report of PatS as a diffusible inhibitor of heterocyst differentiation (73), the specific partners that interact directly in the process of PatS-dependent heterocyst regulation have been difficult to identify, both *in vitro* and *in vivo*. While multiple groups have demonstrated that the RGSGR carboxyl terminus of PatS is capable of inhibiting DNA binding of HetR to a region upstream from its own promoter *in vitro* (39, 65), and a R223W mutant of HetR has been shown to be insensitive to *in vivo* over-expression of both PatS and HetN (45), all of these observations have provided only indirect evidence of binding between HetR and PatS. The specific partners that interact directly during PatS disruption of HetR binding to DNA and the direct molecular-level interactions that cause the loss of R223W HetR sensitivity to PatS have not been unambiguously elucidated with support of experimental data. Specifically with regards to the gel shift experiments, no experimental data has been reported in the literature to distinguish between the two following possibilities: 1) that PatS-5 displaces the DNA upon binding directly to HetR, or 2) that PatS-5 displaces HetR upon binding directly to the DNA. Resolving these issues has been impossible using traditional techniques used to observe direct intermolecular interactions, such as gel shift assays and size exclusion chromatography, because these techniques lack sufficient resolution to detect shifts in the migration of either the 70 kDa HetR dimer or the > 20 kDa DNA fragments upon binding to the PatS-5 pentapeptide, which adds only about 0.6 kDa to the overall complex molecular weight. Moreover, PatS lacks any naturally occurring spectroscopic probe that could be used to detect binding between HetR and PatS-5.

In this manuscript, we report the first unequivocal experimental data that shows direct binding between HetR and PatS-5. Two different and independent techniques corroborated that PatS-5 bound directly to HetR, namely EPR spectroscopy and isothermal titration calorimetry. The conclusion from the EPR experiments was based on the observation that the motion of the MTSL spin-label on the HetR L252C mutant was quenched upon titration with PatS-5 in the absence of any DNA in solution. This observation clearly indicated a direct binding interaction between the mutant spin-labeled

HetR and PatS-5. However, these data did not unambiguously indicate that PatS-5 bound in the immediate vicinity of residue 252, since it is possible that PatS-5 could have bound to HetR in a site remote from residue 252 and caused quenching of spin label motion through an allosteric effect or change in tertiary structure. However, given that *in vivo* experiments indicated that HetR residues 250-256 were critical for sensitivity to PatS-5, it would not be surprising if PatS-5 bound HetR in the C-terminal domain in the vicinity of amino acids 250-256. Besides demonstrating that PatS-5 bound directly to HetR, the DEER EPR experiments also demonstrated that HetR persisted as a dimer even after binding PatS-5. The isothermal titration calorimetry data independently corroborated that PatS-5 bound directly to HetR, using the native HetR protein, and without the mutation or MTSL spin-label required for the EPR experiments. The ITC experiments also firmly established that the stoichiometry of binding was one PatS-5 peptide per HetR monomer, or said differently, two molecules of PatS-5 bind to each HetR dimer. By employing isothermal titration calorimetry, we were not only able to corroborate that PatS-5 bound directly to HetR in the absence of DNA, but we were also able to measure its dissociation constant and determine thermodynamic parameters, which indicated that binding of PatS-5 to HetR was enthalpically-driven.

The high affinity of HetR for the 29 base pair inverted repeat containing upstream *hetP* DNA sequence enabled a semi-quantitative analysis of the stoichiometry of binding of HetR to this DNA fragment, which indicated HetR likely bound to this DNA in a 2:1 ratio, *i.e.*, HetR bound the inverted repeat containing upstream *hetP* DNA sequence as a homodimer. Finally, based on the determination that the shifted species was a 2:1 complex of HetR to DNA, we concluded that the supershifted complex possibly consisted of two HetR homodimers bound to a single *hetP* DNA sequence, which is consistent with the observation that the supershifted species was only detected when the ratio of HetR to DNA was greater than 4:1. Again, despite detecting the supershifted species, the biological relevance of this species, if any, remains to be determined.

The biophysical data were driven by the *in vivo* studies with mutant alleles of *hetR*. The phenotypes of strains with alleles of *hetR* encoding substitutions at residues

R250, E253, E254, L255, and D256 were consistent with decreased sensitivity of the substituted HetR proteins to PatS- and HetN-dependent inhibitory signals. First, both *patS* and *hetN* null mutants as well as the strains with overactive alleles described here differentiated an increased number of heterocysts relative to that of the wild-type, PCC 7120 (14, 73). Second, a *patS* null mutant and the strains described here with overactive alleles differentiated heterocysts in the presence of a fixed source of nitrogen. And third, inactivation of both *patS* and *hetN* simultaneously resulted in the formation of a number of heterocysts in excess of that made by either of the individual mutants (8), similar to the strains with alleles of *hetR* encoding E253D, E254D, L255V, and D256E substitutions.

2.6 Acknowledgements

The authors thank Professor Susan Barnum for providing the genomic DNA of *Anabaena* sp. strain PCC 7120 and Kelly Higa for insightful discussions. Bryan J. Glaenzer, Alisha N. Jones, and Andrea F. Schilling are acknowledged for contributing to early stages of characterization of soluble HetR. This work was supported in part by a grant from NSF (IOS-0919878) to SMC.

2.7 References

1. Adams, D. G. 2000. Heterocyst formation in cyanobacteria. Curr Opin Microbiol 3:618-624.
2. Adams, D. G., and N. G. Carr. 1981. The developmental biology of heterocyst and akinete formation in cyanobacteria. Crit Rev Microbiol 9:45-100.
3. Altenbach, C., T. Marti, H. G. Khorana, and W. L. Hubbell. 1990. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. Science 248:1088-1092.
4. Banks, G. C., B. Mohr, and R. Reeves. 1999. The HMG-I(Y) A.T-hook peptide motif confers DNA-binding specificity to a structured chimeric protein. J Biol Chem 274:16536-16544.

5. Black, T. A., Y. Cai, and C. P. Wolk. 1993. Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in Anabaena. Mol Microbiol 9:77-84.
6. Borbat, P., and J. Freed. 2001. Double quantum ESR and distance measurements. Kluwer, New York, NY.
7. Borbat, P. P., J. H. Davis, S. E. Butcher, and J. H. Freed. 2004. Measurement of large distances in biomolecules using double-quantum filtered refocused electron spin-echoes. J Am Chem Soc 126:7746-7747.
8. Borthakur, P. B., C. C. Orozco, S. S. Young-Robbins, R. Haselkorn, and S. M. Callahan. 2005. Inactivation of patS and hetN causes lethal levels of heterocyst differentiation in the filamentous cyanobacterium Anabaena sp. PCC 7120. Mol Microbiol 57:111-123.
9. Brocks, J. J., G. A. Logan, R. Buick, and R. E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285:1033-1036.
10. Buchaklian, A. H., and C. S. Klug. 2005. Characterization of the Walker A motif of MsbA using site-directed spin labeling electron paramagnetic resonance spectroscopy. Biochemistry 44:5503-5509.
11. Budil, D., S. Lee, S. Saxena, and J. Freed. 1996. Nonlinear-least-squares analysis of slow-motion EPR spectra in one and two dimensions using a modified Levenberg-Marquardt algorithm. Journal of Magnetic Resonance Series a:155-189.
12. Buikema, W. J., and R. Haselkorn. 1991. Characterization of a gene controlling heterocyst differentiation in the cyanobacterium Anabaena 7120. Genes Dev 5:321-330.
13. Calderone, C. T., and D. H. Williams. 2001. An enthalpic component in cooperativity: the relationship between enthalpy, entropy, and noncovalent structure in weak associations. J Am Chem Soc 123:6262-6267.
14. Callahan, S. M., and W. J. Buikema. 2001. The role of HetN in maintenance of the heterocyst pattern in Anabaena sp. PCC 7120. Mol Microbiol 40:941-950.

15. Chiang, Y. W., P. P. Borbat, and J. H. Freed. 2005. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. J Magn Reson 172:279-295.
16. Connelly, P. R., R. A. Aldape, F. J. Bruzzese, S. P. Chambers, M. J. Fitzgibbon, M. A. Fleming, S. Itoh, D. J. Livingston, M. A. Navia, and J. A. Thomson. 1994. Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. Proc Natl Acad Sci U S A 91:1964-1968.
17. Dong, Y., X. Huang, X. Y. Wu, and J. Zhao. 2000. Identification of the active site of HetR protease and its requirement for heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 182:1575-1579.
18. El-Shehawy, R., and D. Kleiner. 2003. The mystique of irreversibility in cyanobacterial heterocyst formation: parallels to differentiation and senescence in eukaryotic cells. Physiologia Plantarum:49-55.
19. Elhai, J., and C. P. Wolk. 1988. Conjugal transfer of DNA to cyanobacteria. Methods Enzymol 167:747-754.
20. Fanucci, G. E., and D. S. Cafiso. 2006. Recent advances and applications of site-directed spin labeling. Curr Opin Struct Biol 16:644-653.
21. Fay, P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev 56:340-373.
22. Feig, A. L. 2007. Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. Biopolymers 87:293-301.
23. Fogg, G. 1944. Growth and heterocyst production in Anabaena cylindrica Lemm. New Phytologist 43:164-175.
24. Fogg, G. 1949. Growth and heterocyst production in Anabaena Cylindrica Lemm. II In relation to carbon and nitrogen metabolism. Annals of Botany 13:241-259.
25. Fritsch, F. 1904. Studies on cyanophyceae. The New Phytologist 3:85-96.
26. Fritsch, F. 1951. The heterocyst: a botanical enigma. Proceedings of the Linnean Society of London 162:194-211.

27. Frías, J. E., E. Flores, and A. Herrero. 1997. Nitrate assimilation gene cluster from the heterocyst-forming cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 179:477-486.
28. Gallon, J. 1992. Reconciling the incompatible: N2 fixation and O2. New Phytologist 122:571-609.
29. Golden, J. W., and H. S. Yoon. 2003. Heterocyst development in Anabaena. Curr Opin Microbiol 6:557-563.
30. Goodwin, G. H., and E. W. Johns. 1973. Isolation and characterisation of two calf-thymus chromatin non-histone proteins with high contents of acidic and basic amino acids. Eur J Biochem 40:215-219.
31. Goodwin, G. H., C. Sanders, and E. W. Johns. 1973. A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. Eur J Biochem 38:14-19.
32. Grosschedl, R., K. Giese, and J. Pagel. 1994. HMG Domain Proteins - Architectural Elements in the Assembly of Nucleoprotein Structures. Trends in Genetics:94-100.
33. Hanson, S. M., D. J. Francis, S. A. Vishnivetskiy, E. A. Kolobova, W. L. Hubbell, C. S. Klug, and V. V. Gurevich. 2006. Differential interaction of spin-labeled arrestin with inactive and active phosphorhodopsin. Proc Natl Acad Sci U S A 103:4900-4905.
34. Harrison, S. C. 1991. A structural taxonomy of DNA-binding domains. Nature 353:715-719.
35. Haselkorn, R. 2008. Cell-cell communication in filamentous cyanobacteria. Mol Microbiol 70:783-785.
36. Haselkorn, R. 1978. Heterocysts. Annual Review of Plant Physiology and Plant Molecular Biology:319-344.
37. Higa, K. C., and S. M. Callahan. 2010. Ectopic expression of hetP can partially bypass the need for hetR in heterocyst differentiation by Anabaena sp. strain PCC 7120. Mol Microbiol 77:562-574.

38. Hoofnagle, A. N., J. W. Stoner, T. Lee, S. S. Eaton, and N. G. Ahn. 2004. Phosphorylation-dependent changes in structure and dynamics in ERK2 detected by SDSL and EPR. Biophys J 86:395-403.
39. Huang, X., Y. Dong, and J. Zhao. 2004. HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. Proc Natl Acad Sci U S A 101:4848-4853.
40. Huffman, J. L., and R. G. Brennan. 2002. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr Opin Struct Biol 12:98-106.
41. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.
42. Huth, J. R., C. A. Bewley, M. S. Nissen, J. N. Evans, R. Reeves, A. M. Gronenborn, and G. M. Clore. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. Nat Struct Biol 4:657-665.
43. Jeschke, G., V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. Timmel, D. Hilger, and H. Jung. 2006. DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. Applied Magnetic Resonance:473-498.
44. Khudyakov, I., and C. P. Wolk. 1997. hetC, a gene coding for a protein similar to bacterial ABC protein exporters, is involved in early regulation of heterocyst differentiation in Anabaena sp. strain PCC 7120. J Bacteriol 179:6971-6978.
45. Khudyakov, I. Y., and J. W. Golden. 2004. Different functions of HetR, a master regulator of heterocyst differentiation in Anabaena sp. PCC 7120, can be separated by mutation. Proc Natl Acad Sci U S A 101:16040-16045.
46. Kim, Y., G. Joachimiak, Z. Ye, T. A. Binkowski, R. Zhang, P. Gornicki, S. M. Callahan, W. R. Hess, R. Haselkorn, and A. Joachimiak. 2011. Structure of transcription factor HetR required for heterocyst differentiation in cyanobacteria. Proc Natl Acad Sci U S A 108:10109-10114.

47. Klug, C. S., S. S. Eaton, G. R. Eaton, and J. B. Feix. 1998. Ligand-induced conformational change in the ferric enterobactin receptor FepA as studied by site-directed spin labeling and time-domain ESR. Biochemistry 37:9016-9023.
48. Klug, C. S., and J. B. Feix. 2008. Methods and applications of site-directed spin labeling EPR spectroscopy. Methods Cell Biol 84:617-658.
49. Kumar, K., R. A. Mella-Herrera, and J. W. Golden. 2010. Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol 2:a000315.
50. Lambein, F., and C. P. Wolk. 1973. Structural studies on the glycolipids from the envelope of the heterocyst of Anabaena cylindrica. Biochemistry 12:791-798.
51. Latchman, D. S. 1997. Transcription factors: an overview. Int J Biochem Cell Biol 29:1305-1312.
52. Limphong, P., M. W. Crowder, B. Bennett, and C. A. Makaroff. 2009. Arabidopsis thaliana GLX2-1 contains a dinuclear metal binding site, but is not a glyoxalase 2. Biochem J 417:323-330.
53. Mchaourab, H. S., T. Kálai, K. Hideg, and W. L. Hubbell. 1999. Motion of spin-labeled side chains in T4 lysozyme: effect of side chain structure. Biochemistry 38:2947-2955.
54. Mchaourab, H. S., M. A. Lietzow, K. Hideg, and W. L. Hubbell. 1996. Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. Biochemistry 35:7692-7704.
55. Meeks, J. C., and J. Elhai. 2002. Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. Microbiol Mol Biol Rev 66:94-121.
56. Milov, A., A. Ponomarev, and Y. Tsvetkov. 1984. Electron electron double-resonance in electron-spin echo - model biradical systems and the sensitized photolysis of decalin. Chemical Physics Letters:67-72.
57. Milov, A., K. Salikhov, and M. Shirov. 1981. Application of ELDOR in electron-spin echo for paramagnetic center space distribution in solids. Fiz Tverd Tela (Leningrad) 23:957-982.

58. Nesmelov, Y. E., R. V. Agafonov, A. R. Burr, R. T. Weber, and D. D. Thomas. 2008. Structure and dynamics of the force-generating domain of myosin probed by multifrequency electron paramagnetic resonance. Biophys J 95:247-256.
59. Nicolaisen, K., A. Hahn, and E. Schleiff. 2009. The cell wall in heterocyst formation by Anabaena sp. PCC 7120. J Basic Microbiol 49:5-24.
60. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-1802.
61. Polyhach, Y., E. Bordignon, and G. Jeschke. 2011. Rotamer libraries of spin labelled cysteines for protein studies. Phys Chem Chem Phys 13:2356-2366.
62. Reeves, R., W. J. Leonard, and M. S. Nissen. 2000. Binding of HMG-I(Y) imparts architectural specificity to a positioned nucleosome on the promoter of the human interleukin-2 receptor alpha gene. Mol Cell Biol 20:4666-4679.
63. Reeves, R., and M. S. Nissen. 1990. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J Biol Chem 265:8573-8582.
64. Risser, D. D., and S. M. Callahan. 2009. Genetic and cytological evidence that heterocyst patterning is regulated by inhibitor gradients that promote activator decay. Proc Natl Acad Sci U S A 106:19884-19888.
65. Risser, D. D., and S. M. Callahan. 2007. Mutagenesis of hetR reveals amino acids necessary for HetR function in the heterocystous cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 189:2460-2467.
66. Schneider, D., and J. Freed. 1989. Calculating slow motional magnetic resonance spectra: A user's guide, p. 1-76. *In* L. Berliner and J. Reuben (ed.), Biological Magnetic Resonance, vol. 8. Plenum, New York, NY.
67. Streicher, W. W., M. M. Lopez, and G. I. Makhatadze. 2009. Annexin I and annexin II N-terminal peptides binding to S100 protein family members: specificity and thermodynamic characterization. Biochemistry 48:2788-2798.

68. Tomitani, A., A. H. Knoll, C. M. Cavanaugh, and T. Ohno. 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proc Natl Acad Sci U S A 103:5442-5447.
69. Williams, D. H., E. Stephens, D. P. O'Brien, and M. Zhou. 2004. Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. Angew Chem Int Ed Engl 43:6596-6616.
70. Wolk, C. P. 1996. Heterocyst formation. Annu Rev Genet 30:59-78.
71. Wolk, C. P., and M. P. Quine. 1975. Formation of one-dimensional patterns by stochastic processes and by filamentous blue-green algae. Dev Biol 46:370-382.
72. Wu, X., D. Liu, M. H. Lee, and J. W. Golden. 2004. patS minigenes inhibit heterocyst development of Anabaena sp. strain PCC 7120. J Bacteriol 186:6422-6429.
73. Yoon, H. S., and J. W. Golden. 1998. Heterocyst pattern formation controlled by a diffusible peptide. Science 282:935-938.
74. Zhang, C. C., S. Laurent, S. Sakr, L. Peng, and S. Bédu. 2006. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. Mol Microbiol 59:367-375.
75. Zhang, J. Y., W. L. Chen, and C. C. Zhang. 2009. hetR and patS, two genes necessary for heterocyst pattern formation, are widespread in filamentous nonheterocyst-forming cyanobacteria. Microbiology 155:1418-1426.
76. Zhang, Z., M. R. Fleissner, D. S. Tipikin, Z. Liang, J. K. Moscicki, K. A. Earle, W. L. Hubbell, and J. H. Freed. 2010. Multifrequency electron spin resonance study of the dynamics of spin labeled T4 lysozyme. J Phys Chem B 114:5503-5521.
77. Zhou, R., Z. Cao, and J. Zhao. 1998. Characterization of HetR protein turnover in Anabaena sp. PCC 7120. Arch Microbiol 169:417-423.

78. Zhou, R., X. Wei, N. Jiang, H. Li, Y. Dong, K. L. Hsi, and J. Zhao. 1998. Evidence that HetR protein is an unusual serine-type protease. Proc Natl Acad Sci U S A 95:4959-4963.

Chapter 3: Differential binding between PatS C-terminal peptide fragments and HetR from *Anabaena* sp PCC 7120

Reproduced with permission from:

Erik A. Feldmann,[1] Shuisong Ni,[1] Indra D. Sahu,[1] Clay H. Mishler,[1] Jeffery D. Levengood,[1] Yegor Kushnir,[1] Robert M. McCarrick,[1] Gary A. Lorigan,[1] Blanton S. Tolbert,[1] Sean M. Callahan,[2] and Michael A. Kennedy*,[1]

[1] *Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056, USA*
2 Department of Microbiology, University of Hawaii, Honolulu, Hawaii 96822, USA

* Corresponding author: Department of Chemistry and Biochemistry, Hughes Laboratories, Room 106, Miami University, 701 High Street, Oxford, Ohio 45056. E-mail: kennedm4@muohio.edu. Phone: 513-529-8267. Fax: 513-529-5715.

This paper has been published in *Biochemistry* 2012, **51**(12):2436-42.

Copyright 2012 American Chemical Society

*Author contributions***:** SN, RMM, GAL, BST, SMC, and MAK contributed to data analysis; IDS, CHM, JDL, and YK contributed to data collection and/or data analysis; EAF contributed to *in vitro* data collection and analysis of recombinantly-expressed HetR proteins including: size-exclusion chromatography, site-directed mutagenesis, MTSL spin labeling, CW EPR spectroscopy, and ITC calorimetry. EAF also wrote the manuscript.

3.1 Abstract

Heterocyst differentiation in the filamentous cyanobacterium *Anabaena* sp. strain PCC 7120 occurs at regular intervals under nitrogen starvation and is regulated by a host of signaling molecules responsive to availability of fixed nitrogen. The heterocyst differentiation inhibitor PatS contains the active pentapeptide RGSGR (PatS-5) at its C-terminus, which is considered the minimum PatS fragment required for normal heterocyst pattern formation. PatS-5 is known to bind HetR, the master regulator of heterocyst differentiation, with a moderate affinity and a sub-micromolar dissociation constant. Here we characterized the affinity of HetR for several PatS C-terminal fragments by measuring the relative ability of each fragment to knockdown HetR binding to DNA in electrophoretic mobility shift assays and using isothermal titration calorimetry (ITC). HetR bound to PatS-6 (ERGSGR) >30 times tighter (K_d = 7 nM) than to PatS-5 (K_d = 227 nM) and >1200 times tighter than to PatS-7 (DERGSGR) (K_d = 9280 nM). No binding was detected between HetR and PatS-8 (CDERGSGR). Quantitative binding constants obtained from ITC measurements were consistent with qualitative results from the gel shift knockdown assays. CW EPR spectroscopy confirmed that PatS-6 bound to a MTSL spin-labeled HetR L252C mutant at a 10-fold lower concentration compared to PatS-5. Substituting the PatS-6 N-terminal glutamate to aspartate, lysine, or glycine did not alter binding affinity, indicating that neither the charge nor size of the N-terminal residue's side-chain played a role in enhanced HetR binding to PatS-6, but rather increased binding affinity resulted from new interactions with the PatS-6 N-terminal residue peptide backbone.

3.2 Introduction

Cyanobacteria are a diverse group of organisms known to exist on Earth for more than 2 billion years (2) and were partially responsible for the oxidation of the planet's atmosphere by photosynthesis (35). In addition, many species of cyanobacteria fix N_2 gas (24), an essential step in the nitrogen cycle, contributing to the pool of organic compounds eventually used in synthesis of nucleic and amino acids as well as numerous

other biomolecules (6). Under nitrogen starved conditions, the filamentous cyanobacterium *Anabaena* sp. strain PCC 7120 differentiates vegetative cells into specialized cells called heterocysts (36), in order to facilitate nitrogen fixation in heterocysts while maintaining photosynthesis in vegetative cells. Differentiation of vegetative cells into heterocysts is restricted to approximately 10% of cells along the filament spaced in a regular pattern (36). Development of heterocysts and maintenance of the pattern appears to be tightly regulated by a network of signaling molecules responsive to the amount of fixed nitrogen available (15, 16, 37).

Diffusible inhibitors of heterocyst differentiation are thought to play an important role in pattern formation via the establishment of inhibitor gradients that promote destruction of positive regulatory factors for heterocyst differentiation (32, 33). The PatS peptide and HetN protein are two inhibitors that contribute to establishing this regulatory gradient in *Anabaena* (28). PatS is a 13 or 17 residue (ambiguity due to two potential ATG start codons (39)) peptide required for initial pattern formation (40) and contains the pentapeptide RGSGR at its C-terminus. PatS has been postulated to control the heterocyst pattern by lateral inhibition (16) via the RGSGR motif, which has been shown to be the minimum sequence required for the normal heterocyst patterned phenotype *in vivo* (39). HetN is a 287 amino acid protein also containing the RGSGR sequence but the expression of *hetN* is not required for initial patterning. Instead, HetN is necessary for proper heterocyst-pattern maintenance during cellular growth and filament elongation afterward (8). However, proteolytic machinery capable of processing PatS or HetN to smaller peptides has yet to be identified. In addition, the active forms of PatS and HetN unfortunately have also not been successfully isolated and are still unknown.

HetR is a 299 amino acid homodimeric protein that acts as a positive signal (4), considered the master regulatory switch for initiation of heterocyst differentiation (27), and has recently been identified as a direct binding partner for the RGSGR carboxyl terminus of PatS (PatS-5) (14). HetR is an autoregulatory transcription factor constitutively expressed at low levels in every cell along the filament (1) and is responsible for activating expression of multiple downstream genes (4). However,

transcription of *hetR* is tightly-regulated and is turned over rapidly in *Anabaena* (42). Transcription of *hetR* is regulated temporally and spatially (1, 27). Under nitrogen starvation, *hetR* transcript levels begin to increase 0.5 h after nitrogen step-down, and continue up to a five-fold increase by 6-18 h (4, 5). Gradients of HetR are established along the filament where a maximum is reached midway between differentiated heterocysts (28), which also predicted to coincide with a minimum in RGSGR concentration (28), and it is here that a vegetative cell is initiated to differentiate into a heterocyst. It has been suggested that these gradients of processed RGSGR-containing peptide forms of PatS and HetN promote the posttranslational degradation of HetR (5) and consequently, play a role in controlling the HetR gradient (28).

There are four transcriptional start positions (tsps) of *hetR* in *Anabaena* at nucleotides -728, -696, -271, and -184 (relative to the translational start site), and correspondingly, each tsp leads to expression of a unique *hetR* mRNA transcript (5). The -184 tsp transcript is involved in maintenance of basal levels of *hetR* under all growth conditions (5, 25). Transcripts starting from the -728 and -696 tsps are also expressed at low levels under nitrogen replete conditions (5, 25) but are significantly up-regulated by NrrA under nitrogen starvation (11, 12). NrrA is the response regulatory protein activated by NtcA (11, 12). NtcA is the global nitrogen regulatory protein that responds to 2-oxoglutarate accumulation when *Anabaena* is nitrogen starved (23) and is also required for expression of *hetR* from -728 tsp (25). The -271 tsp appears to be the only position that requires functional HetR availability (5) and is the only one of the four transcripts that is regulated spatially along the filament (27). HetR was also found to have *in vitro* binding activity with DNA around the -271 tsp (21). Accordingly, the PatS and HetN inhibitors have been shown to have a negative effect on *hetR* transcription from tsp -271, and it has been suggested that PatS may regulate patterning by preventing HetR from activating transcription at this tsp, thus contributing further to establishment of the HetR gradient (27). For these reasons, the -271 tsp is considered the site of HetR autoregulation.

The RGSGR pentapeptide appears to play multiple roles in HetR-related regulation. In addition to promoting HetR destruction, PatS-5 inhibits HetR DNA-binding activity *in vitro* (14, 21, 29). The DNA-binding inhibition appears to be accomplished from tight binding between the peptide and HetR, which is capable of binding two PatS-5 molecules per HetR homodimer. It also appears that the PatS-5 binding site on HetR is in the proximity of the C-terminal amino acid region R250-D256 based on EPR-based cysteine scanning mutagenesis studies (14) guided by *in vivo* observations that conservative mutations of HetR in this region exhibited differential PatS-5 sensitivity (29). The recent report of the HetR crystal structure from *Fischerella* MV11 has also modeled this C-terminal area, among other sites, to be likely candidates for the PatS-5 binding site based on Nest predictions from existing protein-peptide binding surfaces (22). Structures of the HetR-PatS complex or any HetR-DNA complexes, however, remain unknown at this time.

Isothermal titration calorimetry (ITC) has proven to be a valuable technique for studying thermodynamics of protein-protein/peptide/nucleic acid/small molecule interactions, particularly exemplified by the drug discovery industry (13). Such information is valuable when comparing binding interactions of peptide ligands to receptor targets and can be used to provide insight into the binding mechanism (9, 17). When coupled with other biophysical strategies like electron paramagnetic resonance (EPR) spectroscopy, nuclear magnetic resonance spectroscopy, and X-ray crystallography, ITC can be a powerful tool for understanding the fine molecular intricacies of biological processes.

In this report we used ITC and EPR to investigate the *in vitro* HetR binding affinity to several C-terminal PatS fragments of varying length for both the native PatS sequence and for selected amino acid substitutions. We also studied the relative ability of these peptides to disrupt the DNA binding activity of HetR using electrophoretic mobility shift assay (EMSA) knockdown experiments. Surprisingly, our data indicated that HetR binds to PatS-6 with more than 30 times greater affinity compared to PatS-5, which raises

important new questions as to what is the active form of PatS *in vivo*. These surprising results and their implications are discussed in detail below.

3.3 Materials and Methods

3.3.1 Cloning, expression, and purification of HetR protein and preparation of PatS peptide

Cloning, over-expression, and purification of recombinant soluble HetR was performed as described previously (14). DNA binding was assessed using electrophoretic mobility shift assays as described previously (14). The complementary DNA oligonucleotides for the *hetP*-29-mer were synthesized and HPLC purified by Integrated DNA Technologies (Coralville, IA) at 250 nmole scale synthesis for each strand (forward sequence: 5´-GTAGGCGAGGGGTCTAACCCCTCATTACC-3´ and reverse sequence: 5´ GGTAATGAGGGGTTAGACCCCTCGCCTAC-3´). Double stranded *hetP*-29-mer was annealed by suspending equal stoichiometric concentrations of forward and reverse strands together and heating to 85 °C for ten minutes, followed by slow cooling to room temperature. All native and mutant PatS peptides were custom synthesized and purified by Peptide2.0 (Chantilly, VA) on a 5 mg scale, and suspended in 100% Nanopure H_2O to stock 10 mM concentrations. Preparation of agarose gels for EMSA was performed as described previously (14).

3.3.2 EPR spectroscopy

Generation of purified 252C HetR protein (C48A, L252C double mutant) and subsequent site-directed spin labeling with the nitroxide spin radical (1-oxyl-2,2,5,5-tetramethyl-pyrrolin-3-yl)methyl methanethiosulfonate (MTSL) (Toronto Research Chemicals Inc.) was performed as described previously (14). Spin-labeled L252C HetR was concentrated to 200 µM prior to EPR measurements using an Amicon Ultra membrane filter (Millipore). Approximately 30 mL of the HetR and HetR-PatS

complexes were drawn into 1.1 mm internal diameter (1.6 mm external diameter) quartz capillaries. Capillaries were placed into 3 mm internal diameter quartz EPR tubes and inserted into the microwave cavity. CW EPR spectra were collected at the Ohio Advanced EPR Laboratory at X-band on a Bruker EMX CW-EPR spectrometer using an ER041xG microwave bridge and ER4119-HS cavity coupled with a BVT 3000 nitrogen gas temperature controller (temperature stability ±0.2 K). CW EPR spectra were collected by signal averaging 15 42-s field scans (consisting of 1024 points and 40 ms time constants and conversion times) with a center field of 3370 G and sweep width of 100 G, microwave frequency of 9.5 GHz, modulation frequency of 100 kHz, modulation amplitude of 1 G, microwave power of 1 mW at 298 K. EPR spectral simulations to extract best-fit values of the dynamics properties (Table 3.2) were performed as described previously (3, 14, 20, 26, 31, 41).

3.3.3 ITC Calorimetry

ITC data were collected at 25 °C on a VP-ITC titration calorimeter (Microcal, Northampton, MA), and the resulting isotherms were deconvoluted and fitted using the ORIGIN for ITC software package (Microcal, Piscataway, NJ) provided with the instrument. It was necessary to purify a large homogenous batch of HetR in order to make a fair comparison between peptides. Several passages of HetR were made through a gel filtration column (Pharmacia Superdex200 HiLoad size exclusion column), and once enough was collected, the HetR stock was dialyzed overnight into buffer containing 1 M NaCl, 10% (w/v) glycerol, 10 mM Tris, pH 7.8, and 300 mM imidazole. This reference buffer was later used in the reference cell of the ITC titration as well as for diluting and equilibrating concentrated stocks of synthetic PatS peptides for the ITC titrant. All peptides were diluted to a final concentration of 100 µM with the exception of PatS-6nD, which required 300 µM concentration and 15 µM HetR cell concentration for optimal signal:noise sensitivity. All other HetR cell concentrations were 5 µM. The sample cell was loaded with 1.46 mL of degassed HetR, and the injection syringe was loaded with 0.6 mL of PatS peptide. Blank reference titrations were recorded for each

PatS peptide to account for the heats of dilution into sample buffer. Titrations were performed in duplicate and consisted of 35 8 µL injections following an initial 4 µL injection.

3.4 Results

3.4.1 EMSA knockdown comparison of native PatS peptides

HetR has been shown to bind to a 29-base pair inverted-repeat-containing DNA sequence in the promoter region of *hetP*, a gene encoding a protein that is involved in regulating heterocyst differentiation downstream of HetR (14, 18). This 29-mer DNA substrate containing the inverted repeat (5'- GAGGGGTCTAACCCCTC-3´) was recently used to assess the stoichiometry of HetR-DNA/PatS-5 complexes in EMSA experiments (14). Here, we used the *hetP*-29-mer to study disruption of HetR binding to DNA by various length C-terminal PatS fragments. In the absence of added PatS, we observed completely shifted DNA by EMSA at 3:1 ratios of HetR:DNA or greater (Figure 3.1). At or above this ratio, an additional supershifted species was observed for the HetR-*hetP*-29-mer complex. PatS-5 disrupted the supershifted species at a 0.5:1 ratio (PatS:HetR) and completely knocked down the shifted species at a 10:1 ratio (14). Here, we incubated the 4:1 ratio HetR-*hetP*-29-mer complex separately with 1:1 ratios of PatS-5, PatS-6, PatS-7, and PatS-8 (Figure 3.1). We observed qualitatively that PatS-7 had a similar disruptive effect on HetR DNA binding activity as PatS-5. Surprisingly, PatS-6 disrupted all shifted DNA at a 1:1 ratio, whereas PatS-8 failed to disrupt HetR/DNA shifts at all. This result was unexpected based on assumptions from an earlier report that a strain of *Anabaena* expressing the PatS-8 minigene on a plasmid was still capable of inhibiting heterocyst differentiation (38). This result provides indirect evidence that PatS must be processed to a fragment shorter than PatS-8 in order to be active *in vivo*.

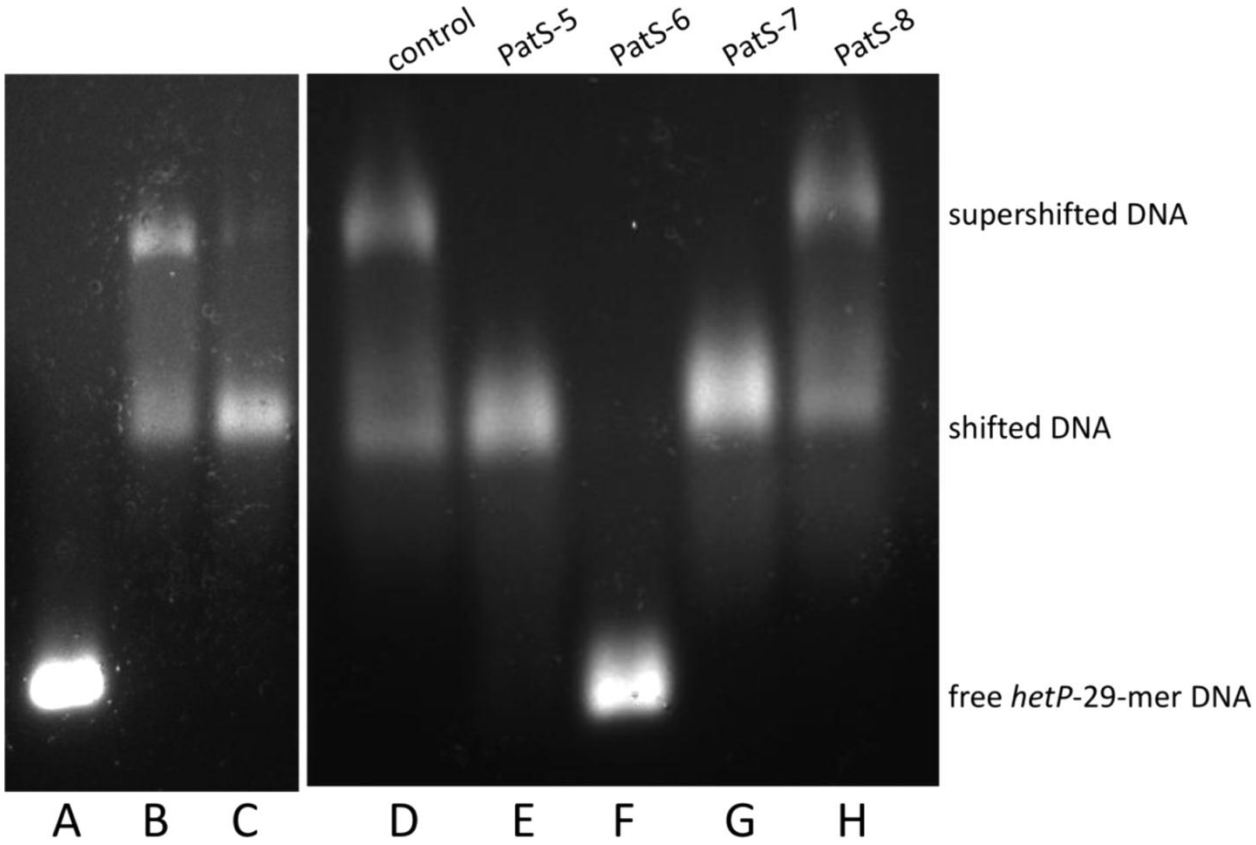

Figure 3.1 EMSA knockdown gel of PatS C-terminal peptides with the HetR-*hetP*-29-mer DNA complex. Lane (A) consists of the *hetP*-29-mer DNA alone, Lane (B) contains a 4:1 HetR: *hetP*-29-mer DNA complex, Lane (C) contains a 3:1 HetR: *hetP*-29-mer DNA complex, Lane (D) consists of the 4:1 HetR-*hetP*-29-mer DNA complex control sample, (E) is the complex incubated with a 1:1 ratio (relative to HetR) of PatS-5, (F) 1:1 incubation with PatS-6, (G) 1:1 incubation with PatS-7, and (H) 1:1 incubation with PatS-8. The bottom band represents free *hetP*-29-mer DNA, the middle band represents the primary shifted species of the HetR-*hetP*-29-mer DNA complex, and the top band represents the supershifted species of the complex.

3.4.2 Thermodynamics and binding affinities of native and mutant PatS peptides binding to HetR

The observation that disruption of HetR binding to DNA depended on the length of the native PatS C-terminal peptides in EMSA experiments prompted us to quantitatively determine the binding affinities between HetR and each peptide using isothermal titration calorimetry (ITC). We have previously reported ITC data for HetR binding to PatS-5 (14). Here, ITC titration experiments were initially performed with PatS-6, PatS-7, and PatS-8 using the same batch of HetR. All resulting ITC data was fitted with a single binding site model, which allowed determination of the binding stoichiometry *n*, association constant K_a, and enthalpy of binding $\Delta H°$ (Figure 3.2 and Table 3.1). The $\Delta G°$ was determined from K_a, the dissociation constant determined from the inverse of K_a, and $\Delta S°$ calculated using $\Delta G°$ and $\Delta H°$ values. Each PatS peptide bound to HetR in a 1:1 ratio with the exception of PatS-8, for which we could not detect any heats of binding, in agreement with the EMSA observations (PatS-8 also had undetectable binding by ITC and EMSA under reducing conditions of dithiothreitol, data not shown). The ITC results for PatS-6 indicated surprisingly tight binding between PatS-6 and HetR (K_d = 7.36+/-0.45 nM), representing a >30-fold increase in affinity compared to PatS-5 (K_d = 227+/-23 nM) (14). Furthermore, a >1200-fold reduction in binding affinity was observed for PatS-7 (K_d = 9.28+/-2.43 µM) compared to PatS-6 (and 40-fold reduction compared to PatS-5). The relative HetR binding affinities to PatS-5 and PatS-7 were not discernable from the EMSA data alone since PatS-7 and PatS-5 appeared to similarly disrupt HetR DNA binding activity (Figure 3.1). The ITC data indicated that HetR binding to PatS-5, PatS-6, and PatS-7 was strongly exothermic for all binding reactions, driven primarily by enthalpic contributions compared with entropic contributions, with the exception of PatS-6 which had comparable magnitudes of binding entropy and binding enthalpy ($\Delta H°$ = -5.85+/-0.03 kcal/mol and $T\Delta S°$ = 5.25+/-0.05 kcal/mol). PatS-6, however, had a larger overall Gibb's free energy ($\Delta G°$ = -11.05+/-0.08 kcal/mol) compared to PatS-5. PatS-7 was the only ligand that exhibited a negative entropic contribution to binding ($T\Delta S°$ = -5.87+/-0.39 kcal/mol) and a larger magnitude in

enthalpy of binding ($\Delta H°$ = -12.77+/-0.23 kcal/mol) compared to the overall Gibb's free energy ($\Delta G°$ = -6.90+/-0.16 kcal/mol).

Having discovered that the N-terminal glutamate residue of PatS-6 conferred more than an order of magnitude tighter binding to HetR compared to PatS-5, we next investigated specifically whether the glutamate residue was necessary for tighter binding, or whether other amino acids in this position would maintain high affinity. To address this question, we designed a series of PatS-6 peptides containing various amino acid substitutions at the N-terminal position to determine whether the charge and/or size of the amino acid side chain was essential for increased binding affinity to HetR. The altered peptides included a conservative substitution with another negatively-charged amino acid, aspartic acid (PatS-6nD, DRGSGR), substitution of glutamate with a positively-charged amino acid, lysine (PatS-6nK, KRGSGR), and substitution with a neutral charged amino acid that lacked a sidechain altogether, glycine (PatS-6nG, GRGSGR). ITC titrations of HetR were repeated with these peptides, and the results are shown in Figure 3.2. Each PatS-6 N-terminal variant bound HetR with a 1:1 stoichiometry and exhibited a highly exergonic Gibb's free energy (Figure 3.2 and Table 3.1). PatS-6nK, with its N-terminal substitution of a long-chain positively-charged lysine, had a comparable binding affinity (K_d = 5.71+/-1.33 nM) and thermodynamic parameters ($\Delta H°$ = -5.85+/-0.04 kcal/mol, $T\Delta S°$ = 5.44+/-0.11 kcal/mol, and $\Delta G°$ = -11.25+/-0.14 kcal/mol) compared to the wild-type PatS-6. Similarly, PatS-6nG also had a comparable binding affinity (K_d = 7.88+/-1.55 nM) and Gibb's free energy ($\Delta G°$ = -11.07+/-0.12 kcal/mol) despite replacing the negatively-charged glutamate with an uncharged glycine that lacked a side-chain entirely. However, the enthalpic contribution for PatS-6nG binding was reduced by nearly 40% ($\Delta H°$ = -3.68+/-0.03 kcal/mol), and the entropic contribution was increased by 40% ($T\Delta S°$ = 7.39+/-0.15 kcal/mol), indicating an entropically- rather than an enthalpically-driven binding event. To our further surprise, PatS-6nD bound to HetR with a ~3.5-fold reduced binding affinity (K_d = 25.7+/-2.6 nM) compared to native PatS-6, despite the conservative substitution with aspartic acid. The Gibb's free energy ($\Delta G°$ = -10.36+/-0.08 *kcal/mol*) remained consistent with the native

PatS-6 and the other PatS-6 variants, yet similarly to the shorter PatS-6nG, the enthalpic contribution was reduced by nearly 40% ($\Delta H°$ = -3.60+/-0.01 kcal/mol) and the entropic contribution was increased by 30% ($T\Delta S°$ = 6.75+/-0.07 kcal/mol), resulting in an entropically-driven interaction.

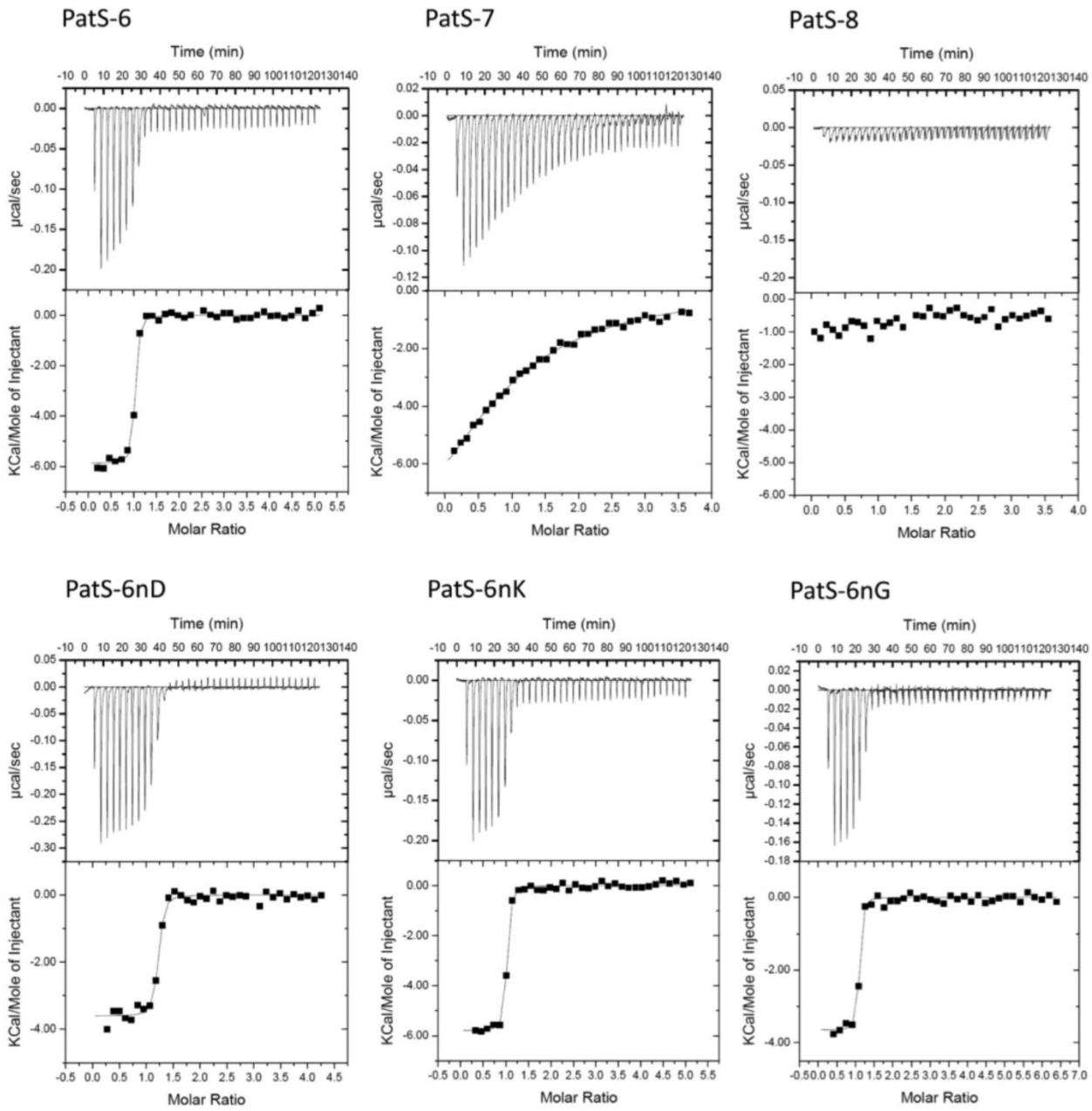

Figure 3.2 Representative raw and fit ITC isotherms for PatS peptides titrated into HetR. Shown at the *top* are the raw data for PatS ligands titrated into HetR solutions and the *bottom* shows the processed binding isotherms calculated using the best-fit parameters for single binding site models. Experiments were performed in duplicate at 25 °C.

peptide	n (no. of sites)	K_d (nM)	$\Delta G°$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$T\Delta S°$ (kcal/mol)
PatS-5*	1.04 ± 0.01*	227 ± 23*	−9.06 ± 0.06*	−8.34 ± 0.41*	0.73 ± 0.46*
PatS-6	0.99 ± 0.02	7.36 ± 0.45	−11.05 ± 0.08	−5.85 ± 0.03	5.25 ± 0.05
PatS-6nD	1.17 ± 0.01	25.7 ± 2.6	−10.36 ± 0.08	−3.60 ± 0.01	6.75 ± 0.07
PatS-6nK	0.97 ± 0.01	5.71 ± 1.33	−11.25 ± 0.14	−5.85 ± 0.04	5.44 ± 0.11
PatS-6nG	1.04 ± 0.01	7.88 ± 1.55	−11.07 ± 0.12	−3.68 ± 0.03	7.39 ± 0.15
PatS-7	0.95 ± 0.02	9280 ± 2430	−6.90 ± 0.16	−12.77 ± 0.23	−5.87 ± 0.39

Table 3.1 Summary of ITC data for PatS C-terminal peptides titrated into HetR solutions. Errors are reported as the standard deviation between measurements for each set of titrations. The asterisk represents data reported previously (14).

3.4.3 EPR spectroscopy of spin-labeled HetR titrated with PatS-6

Continuous wave (CW) EPR was used to study PatS-6 binding to a HetR double mutant (C48A, L252C) spin labeled with the nitroxide free radical probe MTSL on the cysteine at residue 252. We have previously shown that the CW EPR spectrum of this spin label HetR mutant experiences motional averaging that is quenched when HetR binds to PatS-5 (14). Here, similarly as with PatS-5, we observed significant quenching of the conformational dynamics of the electron radical associated with the MTSL nitroxide group attached to Cys252 (Figure 3.3). However, binding to PatS-6 quenched the spin label motion at a 10-fold lower concentration compared to PatS-5 (*i.e.*, at a 1:1 ratio compared with a 10:1 ratio in the previous PatS-5 study (14)). The diffusion rate of the MTSL electron radical in the PatS-6 bound state approached $R_{iso} = 9.12x10^5$ sec^{-1} compared to the free state ($R_{iso} = 5.25x10^6$ sec^{-1}) (simulation parameters are in Table 3.2), indicating almost an order of magnitude reduction in the diffusion rate of the spin label after binding to PatS-6. PatS-8 was titrated to a 10:1 excess into a spin-labeled sample of HetR L252C; however, no quenching of motion was observed (Figure 3.4), serving as another negative control similar to the Poly-G pentapeptide (GGGGG) used previously (14). Taken together, the site-directed spin labeling CW EPR titration data were in strong agreement with the observed EMSA and ITC data, indicating that HetR has a much higher affinity for PatS-6 compared to PatS-5.

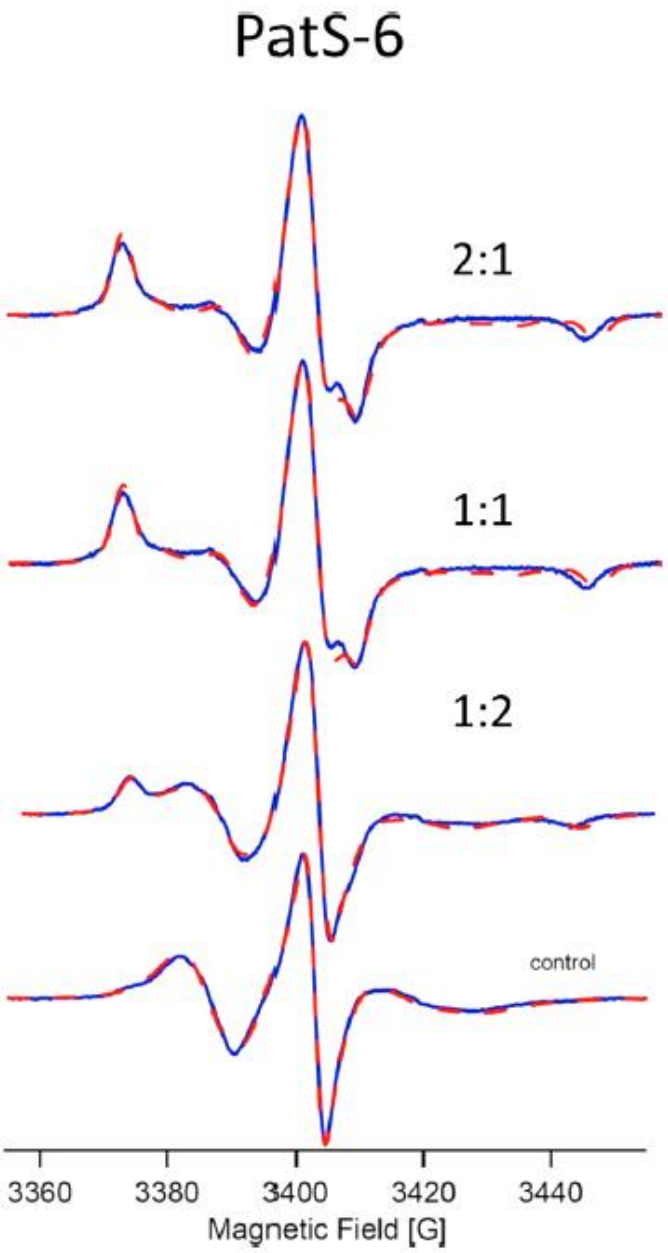

Figure 3.3 CW EPR spectra of PatS-6 titrations into the HetR L252C mutant spin labeled with MTSL. The control spectrum of HetR at 200 µM concentration is shown at the bottom. The remaining spectra consist of PatS-6 titrated into the HetR solution at increasing stoichiometric ratios (indicated above each spectrum). CW X-band EPR spectra were recorded at room temperature. Experimental data are represented by the solid blue lines and best-fit data simulations are represented by the red dashed lines.

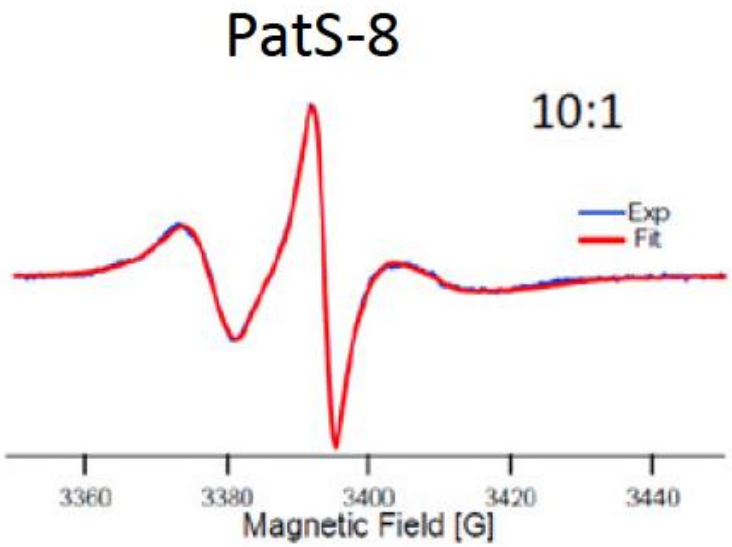

Figure 3.4 CW X-band EPR spectrum of PatS-8 titrated into the HetR L252C mutant spin labeled with MTSL collected at room temperature. Shown above is the endpoint of the titration where the concentration of PatS-8 in solution was 2 mM and the HetR concentration was 200 µM with the PatS-8 to HetR ratio labeled above the spectrum. Experimental data are represented as a dashed blue line, and the best-fit simulation is represented as a solid red line.

sample		A_{xx} (G)	A_{yy} (G)	A_{zz} (G)	g_{xx}	g_{yy}	g_{zz}	log(R_{xx})	log(R_{yy})	log(R_{zz})	log(R_{iso})	β (deg)	% site 1	% site 2
control		7.24	5.94	35.55	2.0083	2.0049	2.0023	7.49	8.43	4.25	6.72	27.67	100.00	0.00
+PatS-6 1:2	site 1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.83	8.53	4.20	6.52	22.48	52.11	47.89
	site 2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.63	6.17	7.64	6.15	2.27		
+PatS-6 1:1	site 1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.88	8.54	4.25	6.56	22.49	18.62	81.38
	site 2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.59	6.06	7.59	6.08	2.28		
+PatS-6 2:1	site 1	7.24	5.94	35.55	2.0083	2.0049	2.0023	6.88	8.54	4.25	6.56	22.49	15.12	84.88
	site 2	7.24	5.94	35.55	2.0083	2.0049	2.0023	4.22	6.07	7.60	5.96	2.28		

Table 3.2 Simulation parameters for CW X-band EPR data of the PatS-6 titration into a solution of HetR L252C mutant spin-labeled with MTSL.

3.5 Discussion

While the functional forms of PatS and HetN that regulate formation of patterns of heterocysts in filamentous cyanobacteria remain unknown, PatS-5 has served as a surrogate for the full-length PatS peptide since its discovery as the minimally required PatS peptide fragment for normal patterns of heterocyst differentiation, and recent evidence suggests that the PatS-5 sequence found in HetN is also required for HetN function (19, 28). However, here we have shown that PatS-6 binds HetR with a 30-fold greater affinity compared to PatS-5, suggesting that the interaction of HetR with PatS-6 may be the more efficient of the two and raising the alternative possibility that PatS-6 might be the active processed form of PatS *in vivo* instead of PatS-5. Our results clearly indicated that the length of the PatS sequence, in this case PatS-5 (RGSGR), PatS-6 (ERGSGR), PatS-7 (DERGSGR), or PatS-8 (CDERGSGR), has the potential to play a critical role in regulating heterocyst differentiation inhibition based on observed differences in their binding affinities to HetR. Since the active forms of PatS or HetN have not, to date, been successfully isolated, it is plausible that any one of these inhibitory sequences may accumulate as a result of protease processing machinery, and regulate heterocyst differentiation via the aforementioned gradient of inhibitors that promote destruction of HetR and other positive regulatory factors. However, processing at least to PatS-7 sequences or shorter appear to be minimally required to enable a binding interaction between HetR and PatS fragments. The biological significance of these findings remains to be shown, and testing these hypotheses will require further *in vivo* studies.

After the discovery that HetR has significantly higher affinity for PatS-6 compared to PatS-5, we set out to understand the biophysical factors that contribute to the tighter binding. One obvious question that needed to be answered was whether or not the naturally occurring N-terminal glutamate in PatS-6 was special or required for the increased binding affinity. We had already been able to elucidate important information about the PatS-HetR binding interaction from the ITC studies for HetR binding to the native PatS peptide sequences. It was clear from the thermodynamics information

derived from the ITC data that all HetR-PatS peptide binding interactions were highly exothermic based on calculated $\Delta H°$ values. The relatively large negative enthalpy and positive entropy associated with HetR-PatS peptide-binding interactions was consistent with ionic interactions and/or formations of hydrogen bonds (7, 10, 17, 30, 34) in the complexes, which would be expected based on the charged nature of the PatS ligands tested at pH 7.8. In order to explicitly probe the importance of the N-terminal glutamate in PatS-6, three modified PatS-6 sequences were prepared with strategically selected substitutions of the N-terminal glutamate, and the binding affinities were measured by ITC and compared to the native PatS-6 peptide. Surprisingly, replacing the native N-terminal glutamate of PatS-6 with a long chain positively-charged lysine did not compromise the tight binding affinity of HetR, indicating that the negatively-charged glutamate carboxyl side chain does not make an important contribution to the tight binding observed between PatS-6 and HetR. Since the charge on the N-terminal amino acid side chain did not contribute to increased binding affinity, this left the possibility that the increased affinity was due either to a new hydrophobic interaction, likely a hydrogen bond, between HetR and the additional peptide backbone formed between the N-terminal residue and the first arginine residue in the RGSGR segment. These possibilities were clearly addressed using data collected for a PatS-6 sequence that contained a substitution of the N-terminal glutamate with a neutral-charged amino acid that lacked a side chain altogether, namely the glycine residue in PatS-6nG. The ITC data for the PatS-6nG indicated that HetR bound to this peptide without any loss in binding affinity compared to the native PatS-6 sequence. This result indicated that the length and/or nature of the side chain was not important for binding to HetR, ruling out the importance of a new hydrophobic interaction. Rather, the data indicated that the N-terminal residue confers increased binding affinity with HetR purely through a new peptide backbone interaction, possibly through the formation of new hydrogen bonds. While the ITC data on the various length native PatS peptide fragments and on the PatS-6 peptides that contained various N-terminal amino acid substitutions have shed important light regarding the specific interactions between HetR and PatS, the detailed molecular level understanding

of the relative binding affinities of the PatS peptides will only be fully realized once the three-dimensional structures of the complexes become available.

3.6 Acknowledgements

The authors thank Professor Susan Barnum for providing the genomic DNA of *Anabaena* sp. strain PCC 7120 and Kelly Higa and Patrick Videau for insightful discussions. This work was supported in part by a grant from NSF (IOS-0919878) to SMC.

3.7 References

1. Black, T. A., Y. Cai, and C. P. Wolk. 1993. Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in Anabaena. Mol Microbiol 9:77-84.
2. Brocks, J. J., G. A. Logan, R. Buick, and R. E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285:1033-1036.
3. Budil, D., S. Lee, S. Saxena, and J. Freed. 1996. Nonlinear-least-squares analysis of slow-motion EPR spectra in one and two dimensions using a modified Levenberg-Marquardt algorithm. Journal of Magnetic Resonance Series a:155-189.
4. Buikema, W. J., and R. Haselkorn. 1991. Characterization of a gene controlling heterocyst differentiation in the cyanobacterium Anabaena 7120. Genes Dev 5:321-330.
5. Buikema, W. J., and R. Haselkorn. 2001. Expression of the Anabaena hetR gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. Proc Natl Acad Sci U S A 98:2729-2734.
6. Burris, R. H., and G. P. Roberts. 1993. Biological nitrogen fixation. Annu Rev Nutr 13:317-335.

7. Calderone, C. T., and D. H. Williams. 2001. An enthalpic component in cooperativity: the relationship between enthalpy, entropy, and noncovalent structure in weak associations. J Am Chem Soc 123:6262-6267.
8. Callahan, S. M., and W. J. Buikema. 2001. The role of HetN in maintenance of the heterocyst pattern in Anabaena sp. PCC 7120. Mol Microbiol 40:941-950.
9. Choi, H. J., J. C. Gross, S. Pokutta, and W. I. Weis. 2009. Interactions of plakoglobin and beta-catenin with desmosomal cadherins: basis of selective exclusion of alpha- and beta-catenin from desmosomes. J Biol Chem 284:31776-31788.
10. Connelly, P. R., R. A. Aldape, F. J. Bruzzese, S. P. Chambers, M. J. Fitzgibbon, M. A. Fleming, S. Itoh, D. J. Livingston, M. A. Navia, and J. A. Thomson. 1994. Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. Proc Natl Acad Sci U S A 91:1964-1968.
11. Ehira, S., and M. Ohmori. 2006. NrrA directly regulates expression of hetR during heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 188:8520-8525.
12. Ehira, S., and M. Ohmori. 2006. NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium Anabaena sp. strain PCC 7120. Mol Microbiol 59:1692-1703.
13. Falconer, R. J., A. Penkova, I. Jelesarov, and B. M. Collins. 2010. Survey of the year 2008: applications of isothermal titration calorimetry. J Mol Recognit 23:395-413.
14. Feldmann, E. A., S. Ni, I. D. Sahu, C. H. Mishler, D. D. Risser, J. L. Murakami, S. K. Tom, R. M. McCarrick, G. A. Lorigan, B. S. Tolbert, S. M. Callahan, and M. A. Kennedy. 2011. Evidence for Direct Binding between HetR from Anabaena sp. PCC 7120 and PatS-5. Biochemistry 50:9212-9224.
15. Flores, E., and A. Herrero. 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. Nat Rev Microbiol 8:39-50.

16. Golden, J. W., and H. S. Yoon. 2003. Heterocyst development in Anabaena. Curr Opin Microbiol 6:557-563.
17. Grässlin, A., C. Amoreira, K. K. Baldridge, and J. A. Robinson. 2009. Thermodynamic and computational studies on the binding of p53-derived peptides and peptidomimetic inhibitors to HDM2. Chembiochem 10:1360-1368.
18. Higa, K. C., and S. M. Callahan. 2010. Ectopic expression of hetP can partially bypass the need for hetR in heterocyst differentiation by Anabaena sp. strain PCC 7120. Mol Microbiol 77:562-574.
19. Higa, K. C., R. Rajagopalan, D. D. Risser, O. S. Rivers, S. K. Tom, P. Videau, and S. M. Callahan. 2012. The RGSGR amino acid motif of the intercellular signalling protein, HetN, is required for patterning of heterocysts in Anabaena sp. strain PCC 7120. Mol Microbiol 83:682-693.
20. Hoofnagle, A. N., J. W. Stoner, T. Lee, S. S. Eaton, and N. G. Ahn. 2004. Phosphorylation-dependent changes in structure and dynamics in ERK2 detected by SDSL and EPR. Biophys J 86:395-403.
21. Huang, X., Y. Dong, and J. Zhao. 2004. HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. Proc Natl Acad Sci U S A 101:4848-4853.
22. Kim, Y., G. Joachimiak, Z. Ye, T. A. Binkowski, R. Zhang, P. Gornicki, S. M. Callahan, W. R. Hess, R. Haselkorn, and A. Joachimiak. 2011. Structure of transcription factor HetR required for heterocyst differentiation in cyanobacteria. Proc Natl Acad Sci U S A 108:10109-10114.
23. Laurent, S., H. Chen, S. Bédu, F. Ziarelli, L. Peng, and C. C. Zhang. 2005. Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in Anabaena sp. PCC 7120. Proc Natl Acad Sci U S A 102:9907-9912.
24. Meeks, J. C., C. P. Wolk, W. Lockau, N. Schilling, P. W. Shaffer, and W. S. Chien. 1978. Pathways of assimilation of [13N]N2 and 13NH4+ by cyanobacteria with and without heterocysts. J Bacteriol 134:125-130.

25. Muro-Pastor, A. M., A. Valladares, E. Flores, and A. Herrero. 2002. Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. Mol Microbiol 44:1377-1385.
26. Nesmelov, Y. E., R. V. Agafonov, A. R. Burr, R. T. Weber, and D. D. Thomas. 2008. Structure and dynamics of the force-generating domain of myosin probed by multifrequency electron paramagnetic resonance. Biophys J 95:247-256.
27. Rajagopalan, R., and S. M. Callahan. 2010. Temporal and spatial regulation of the four transcription start sites of hetR from Anabaena sp. strain PCC 7120. J Bacteriol 192:1088-1096.
28. Risser, D. D., and S. M. Callahan. 2009. Genetic and cytological evidence that heterocyst patterning is regulated by inhibitor gradients that promote activator decay. Proc Natl Acad Sci U S A 106:19884-19888.
29. Risser, D. D., and S. M. Callahan. 2007. Mutagenesis of hetR reveals amino acids necessary for HetR function in the heterocystous cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 189:2460-2467.
30. Ross, P. D., and S. Subramanian. 1981. Thermodynamics of protein association reactions: forces contributing to stability. Biochemistry 20:3096-3102.
31. Schneider, D., and J. Freed. 1989. Calculating slow motional magnetic resonance spectra: A user's guide, p. 1-76. *In* L. Berliner and J. Reuben (ed.), Biological Magnetic Resonance, vol. 8. Plenum, New York, NY.
32. Wilcox, M., G. J. Mitchison, and R. J. Smith. 1973. Pattern formation in the blue-green alga Anabaena. II. Controlled proheterocyst regression. J Cell Sci 13:637-649.
33. Wilcox, M., G. J. Mitchison, and R. J. Smith. 1973. Pattern formation in the blue-green alga, Anabaena. I. Basic mechanisms. J Cell Sci 12:707-723.
34. Williams, D. H., E. Stephens, D. P. O'Brien, and M. Zhou. 2004. Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from

ligand-induced reductions in motion within receptors and enzymes. Angew Chem Int Ed Engl 43:6596-6616.

35. Wolk, C. P. 1996. Heterocyst formation. Annu Rev Genet 30:59-78.
36. Wolk, C. P., A. Ernst, and J. Elhai. 1994. Heterocyst metabolism and development, vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.
37. Wolk, C. P., and M. P. Quine. 1975. Formation of one-dimensional patterns by stochastic processes and by filamentous blue-green algae. Dev Biol 46:370-382.
38. Wu, X., D. Liu, M. H. Lee, and J. W. Golden. 2004. patS minigenes inhibit heterocyst development of Anabaena sp. strain PCC 7120. J Bacteriol 186:6422-6429.
39. Yoon, H. S., and J. W. Golden. 1998. Heterocyst pattern formation controlled by a diffusible peptide. Science 282:935-938.
40. Yoon, H. S., and J. W. Golden. 2001. PatS and products of nitrogen fixation control heterocyst pattern. J Bacteriol 183:2605-2613.
41. Zhang, Z., M. R. Fleissner, D. S. Tipikin, Z. Liang, J. K. Moscicki, K. A. Earle, W. L. Hubbell, and J. H. Freed. 2010. Multifrequency electron spin resonance study of the dynamics of spin labeled T4 lysozyme. J Phys Chem B 114:5503-5521.
42. Zhou, R., Z. Cao, and J. Zhao. 1998. Characterization of HetR protein turnover in Anabaena sp. PCC 7120. Arch Microbiol 169:417-423.

Chapter 4: Structural biology of bacterial proteins from the Northeast Structural Genomics Consortium

Sections 4.2 has been accepted for publication in *Journal of Structural and Functional Genomics*

Solution NMR and X-ray crystal structure of *Pseudomonas syringae* Pspto_3016 from protein domain family PF04237 (DUF419) adopt a "double wing" DNA binding motif

Erik A. Feldmann,[1] Jayaraman Seetharaman,[2] Theresa A. Ramelot,[1] Scott Lew,[2] Li Zhao,[3] Keith Hamilton,[3] Colleen Ciccosanti,[3] Rong Xiao,[3] Thomas B. Acton,[3] John K. Everett,[3] Liang Tong,[2] Gaetano T. Montelione,[3,4] Michael A. Kennedy[1*]

[1] Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056 and Northeast Structural Genomics Consortium

[2] Department of Biological Sciences, Columbia University, New York, New York 10027 and Northeast Structural Genomics Consortium

[3] Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854 and Northeast Structural Genomics Consortium

[4] Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

*Corresponding author: Corresponding author: Department of Chemistry and Biochemistry, Hughes Laboratories, Room 106, Miami University, 701 High Street, Oxford, Ohio 45056. E-mail: kennedm4@muohio.edu. Phone: 513-529-8267. Fax: 513-529-5715.

*Author contributions***:** JS and SL contributed to X-ray crystallography; TAR contributed to solution NMR spectroscopy and structure determination; LZ, KH, CC, RX, TBA, JKE, LT, GTM, and MAK contributed to sample preparation and/or read the manuscript; EAF contributed to solution NMR spectroscopy, structure determination, functional characterization, and wrote the manuscript.

Section 4.3 has been published in *Proteins: Structure, Function and Bioinformatics*

Solution NMR structure of Asl3597 from Nostoc sp. PCC7120, the first structure from protein domain family PF12095, reveals a novel fold

Erik A. Feldmann,[1] Theresa A. Ramelot,[1] Yunhuang Yang,[1] Rong Xiao,[2] Thomas B. Acton,[2] John K. Everett,[2] Gaetano T. Montelione,[2,3] Michael A. Kennedy[1*]

[1] Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056 and Northeast Structural Genomics Consortium

[2] Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854 and Northeast Structural Genomics Consortium

[3] Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

*Corresponding author: Corresponding author: Department of Chemistry and Biochemistry, Hughes Laboratories, Room 106, Miami University, 701 High Street, Oxford, Ohio 45056. E-mail: kennedm4@muohio.edu. Phone: 513-529-8267. Fax: 513-529-5715.

Copyright 2012 John Wiley and Sons

*Author contributions***:** TAR and YY contributed to solution NMR spectroscopy and structure determination; RX, TBA, JKE, GTM, and MAK contributed to sample preparation and/or read the manuscript; EAF contributed to solution NMR spectroscopy, structure determination, functional characterization, and wrote the manuscript.

Sections 4.4 contains data to be submitted for publication to *Proteins: Structure Function and Bioinformatics*

Sections 4.5 contains data to be submitted for publication to *Journal of Structural and Functional Genomics*

4.1 Abstract

In this chapter, a collective summary of structural biology studies on four different bacterial proteins, selected by the Northeast Structural Genomics Consortium, is presented in detail from the phases of gene target cloning, to structure determination and functional interpretation. The targets discussed are: (1) the Pspto_3016 uncharacterized protein from the plant pathogen *Pseudomonas syringae*, a putative DNA-binding protein that adopts a "double-wing" DNA binding motif, (2) the hypothetical uncharacterized Asl3597 protein from *Nostoc* sp. strain PCC 7120, which exhibits sequence similarity with the chlororespiratory reduction 7 (CRR7) protein from the chloroplast stroma of *Arabidopsis thaliana*, (3) the uncharacterized protein, Asr4154 of *Nostoc* sp. PCC 7120, a family member of proteins annotated as the 57 kD subunit of the light-independent protochlorophyllide reductase enzymes, and lastly (4) the uncharacterized mucin-binding domain of protein LBA1460 from the bacterium *Lactobacillus acidophilus*.

4.2 Introduction to NESG target ID: PsR293

We have determined the solution NMR (PDB ID 2KFP) and X-ray crystal structures (PDB ID 3H9X) of the 117-residue protein Pspto_3016 (NESG Target ID PsR293) from the plant pathogen *Pseudomonas syringae* (UniProtKB/TrEMBL ID Q880Y4_PSESM), a member of the protein domain family Pfam (7) PF04237 (DUF419). *P. syringae* is a well-known plant pathogen capable of infecting several species ranging from *Arabidopsis thaliana* to the tomato plant (11). The Pspto_3016 protein from *P. syringae* was selected by the Northeast Structural Genomics Consortium (NESG) as part of the National Institutes of Health Protein Structure Initiative (PSI-2 and PSI-Biology) programs. The goals of PSI-2/PSI-Biology include the "structural coverage" of broadly conserved protein domain families by determination of one or more structures from each of several hundred domain families using NMR and X-ray crystallography (16, 43). The Pfam family PF04237 remains an uncharacterized family of bacterial proteins lacking functional annotation. Pfam family PF04237 also includes domains of proteins that have been identified as human bacterial gut metagenomic targets of the PSI-Biology program

(16). Protein YjbR from *Escherichia coli*, (NESG target ID ER226; PDB ID 2FKI) and protein DR2400 from *Deinococcus radiodurans*, (PDB ID 2A1V), for which 3D structures have been determined by solution NMR (62) and X-ray crystallography (U.A. Ramagopal, Y.V. Patskovsky, S.C. Almo, New York Structural GenomiX Research Consortium. Crystal structure of *Deinococcus radiodurans* protein DR2400, Pfam domain DUF419), respectively, are also members of Pfam PF04237. Based on structural similarities to the C-terminal DNA binding domain (21) of the MotA transcription factor (MotCF) from T4 bacteriophage, outlined below, we propose that Pspto_3016 functions as a sequence-specific DNA-binding protein.

4.2.1 Materials and methods

4.2.1.1 Expression, purification, and characterization of Pspto_3016 protein

The *pspto_3016* gene encoding full length Pspto_3016 protein (residues 1-117) from *Pseudomonas syringae* was cloned into a pET21-23C vector with an additional C-terminal His_6 tag (LEHHHHHH) (NESG plasmid ID, PsR293-21.1). The expression vector plasmid for Pspto_3016 (PsR293-21) is available from the PSI Material Repository (http://psimr.asu.edu/) and from the Structural Biology Knowledgebase (26). Pspto_3016 protein was expressed and purified using standard protocols of the NESG consortium in order to prepare [*U*-^{13}C, ^{15}N] and *U*-^{15}N, 5% biosynthetically directed ^{13}C (NC5) samples (1). To summarize, [*U*-^{13}C, ^{15}N] and NC5 Pspto_3016 were expressed in *E. coli* strain BL21 (DE3) + pMGK, grown in MJ9 medium (35) containing (*U*-$^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. Cells were grown at 37 °C with vigorous shaking until mid-log phase growth then induced with 1 mM isopropyl-β-*D*-thiogalactopyranoside (IPTG) and shaken at 17 °C overnight for protein expression. The cells were harvested by centrifugation, lysed by sonication, and the soluble fraction recovered by high speed centrifugation according to standard NESG

protocols (1). Pspto_3016 protein was first purified with a Ni-affinity column (HisTrap HP, 5 mL), followed by additional purification through a gel filtration column (HiLoad 26/60 Superdex 75) using an ÄKTAxpress™ system (GE Healthcare). The buffer used for Ni-affinity chromatography consisted of 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 40 mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), followed by elution in identical buffer with 500 mM imidazole. The protein was then exchanged into buffer for gel-filtration chromatography containing 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM ammonium acetate, pH 4.5. Purified Pspto_3016 protein was subjected to SDS PAGE to assess sample purity (>95%) and MALDI-TOF mass spectrometry to verify molecular mass (15.6 kDa for [*U*-^{13}C, ^{15}N] and 15.0 kDa for NC5 samples). Samples were then concentrated for NMR by centrifugation to a final concentration of 1.1 mM, and then the addition of 10% v/v D_2O to the final solution. Pspto_3016 behaved as a monomeric protein (98% monodispersity) according to static-light scattering measurements in-line with gel filtration using a Shodex KW-802.5 analytical gel filtration column (30 µL injection) coupled to a miniDAWN TREOS detector (Wyatt technologies) for light scattering at λ=690 nm, using a flow rate of 0.5 mL/min at 40 °C. An isotropic overall rotational correlation time (τ_c) of 11.2 ns was estimated from ^{15}N T_1 and $T_{1\rho}$ relaxation data (Supplementary Figure S1), under the conditions used for NMR (100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM ammonium acetate pH 4.5 at 20 °C) consistent with Pspto_3016 behaving in solution as a monomer.

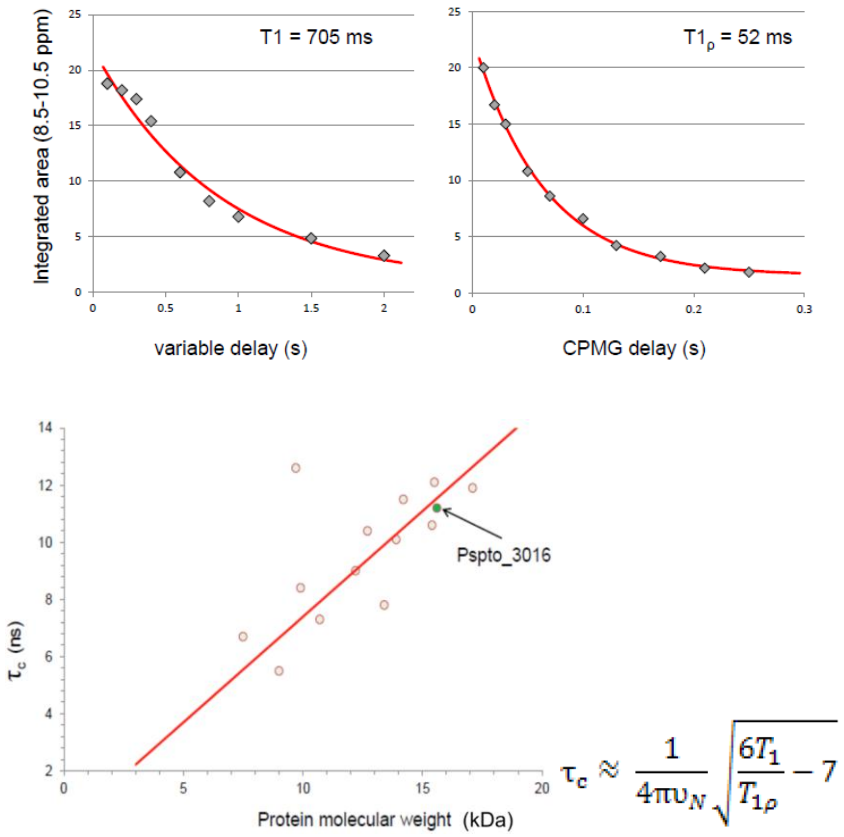

Figure 4.1 1D ^{15}N relaxation data for [*U*-^{13}C, ^{15}N] Pspto_3016 acquired on a Varian Inova 600 MHz spectrometer at 20 °C. (Top) Estimated T_1 and $T_{1\rho}$ relaxation times were obtained by integration from 8.5 to 10.5 ppm (using the Varian intav macro) of the 1D ^{15}N-edited T_1 and $T_{1\rho}$ (CPMG) experiments and fit to an exponential decay (using the Varian t1a or t2a macro, respectively). Longitudinal T_1 relaxation delays were 100, 200, 300, 400, 600, 800, 1000, 1500, and 2000 ms; transverse $T_{1\rho}$ relaxation CPMG delays were 10, 20, 30, 50, 70, 100, 130, 170, 210, and 250 ms; both experiments had 1.5 s recycle delays. (Bottom) Plot of average rotational correlation time, τ_c (ns), vs. protein molecular weight (kDa) of known monomeric proteins solved by solution NMR in the NESG consortium. The τ_c was approximated using the following simplified equation derived from the literature (39), where υ_N is the ^{15}N frequency in Hz. The τ_c of 11.2 ns for Pspto_3016 was obtained using the τ_c Varian macro and is indicated on the plot with an arrow.

4.2.1.2 Protein NMR studies

NMR data were collected at 20 °C on [*U*-^{13}C, ^{15}N] and NC5 Pspto_3016 samples of approximately 270 µL in 5 mm Shigemi NMR tubes on 600 MHz Varian Inova and 850 MHz Bruker Avance III NMR spectrometers equipped with conventional 5 mm probes. The D_2O-exchanged samples were prepared by freezing [*U*-^{13}C, ^{15}N]-Pspto_3016, followed by overnight lyophilization and resuspension into 99.9% D_2O (Acros Organics). Spectra were processed with NMRPipe (15) and visualized with Sparky 3.110 (27). Peptide backbone and side chain resonance assignments were made using the following experiments on [*U*-^{13}C, ^{15}N]-Pspto_3016 samples: ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) and ^{1}H-^{13}C HSQC, HNCA, HNCO, HN(CO)CA, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(C)CH-TOCSY, H(C)CH-COSY, H(CC)(CO)NH-TOCSY, (H)CC(CO)NH-TOCSY, (H)CCH-TOCSY, ^{15}N-edited NOESY-HSQC (τ_m = 70 ms), and two ^{13}C-edited NOESY-HSQC (τ_m = 70 ms) experiments optimized for either aliphatic or aromatic carbons. An additional 4D NOESY (^{13}C-^{13}C-HMQC-NOESY-HMQC) (τ_m = 70 ms) experiment was collected on a D_2O-exchanged [*U*-^{13}C, ^{15}N]-Pspto_3016 sample in order to assign additional NOEs. Stereospecific assignments of isopropyl side chain methyl groups of Val and Leu residues were determined using a high resolution ^{13}C-HSQC constant time experiment collected on the NC5-Pspto_3016 sample. Varian pulse sequences for 2D and 3D experiments came from the Varian BioPack library and the 4D NOESY came from Lewis Kay (University of Toronto). Bruker pulse sequences came from the TopSpin standard library. The amide backbone $^{1}H^{N}$, $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$ and side chain $^{13}C^{\beta}$ and $^{1}H^{\beta}$ resonances were assigned after manual peak picking followed by data submission to AutoAssign-2.3.0 (46) for automatic assignment, while the side chain resonance assignments were completed manually. Final chemical shifts, noesy peak lists, and raw FIDS were deposited in the BioMagResDB (BMRB accession number, 16186). The 2D ^{1}H-^{15}N HSQC spectrum of 1.1 mM [*U*-^{13}C, ^{15}N]-Pspto_3016 with assigned backbone and side chain resonances is shown in Supplementary Figure S2.

The solution NMR structures of Pspto_3016 were calculated with AutoStructure 2.2.1 (32) and CYANA 2.1 (28, 30) to assign NOEs automatically using a combination of assigned resonances, dihedral angle constraints for φ and ψ dihedral angles derived using Talos (14), and NOESY peak lists with peak intensities in Sparky format. Constraints including NOE distances, dihedral angles from Talos (14), and manually derived hydrogen bonds, were converted with PdbStat (6) into XPLOR/CNS format with an increase of 10% to the upper bounds for NOEs. Final structures were calculated with XPLOR-NIH-2.20 (59) and refined in explicit water with CNS version 1.1 (10). Refinement of assigned resonances and NOESY peak lists was guided by NMR RPF (33) quality assessment scores used to assess the goodness-of-fit between the NOESY peak lists and the Xplor-calculated structures. Statistics for calculated structures were used to periodically identify consistent or flagrant NOE constraint violations throughout refinement, before finalizing the NOE distance constraint list. Final constraints and the ensemble of 20 structures sorted by lowest energy were submitted to the Protein Data Bank (PDB ID 2KFP).

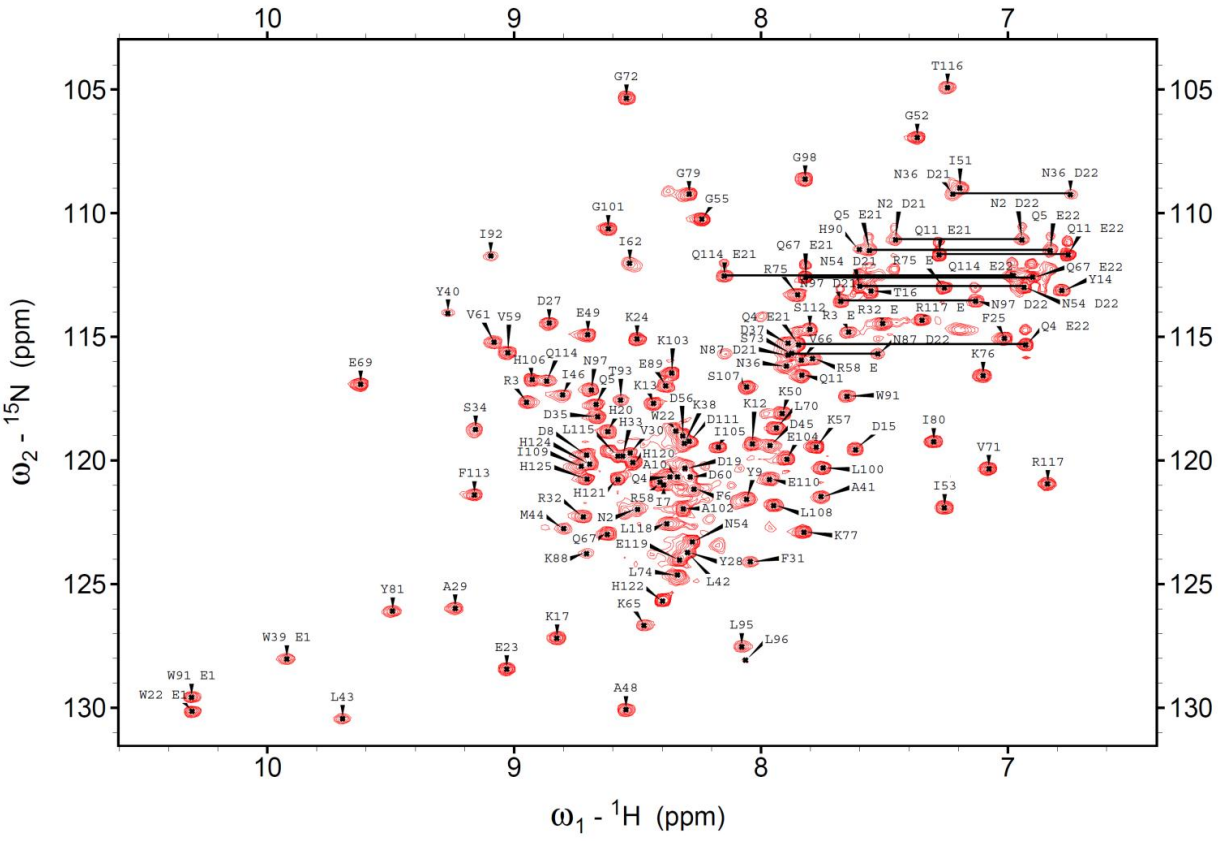

Figure 4.2 Two dimensional ^{1}H-^{15}N HSQC spectrum of a 1.1 mM [*U*-^{13}C, ^{15}N]-Pspto_3016 sample in a 10% v/v D_2O solution of 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM ammonium acetate, pH 4.5, acquired on a Varian Inova 600 MHz spectrometer at 20 °C. Backbone resonance NH assignments are indicated above with their one-letter amino acid codes followed by their sequence number. Also labeled are the assigned side chain NH_2 resonances from Asn and Gln residues and the NH from Arg and Trp residues.

4.2.1.3 Protein crystallization and crystallography

Triclinic crystals of Se-Met-labeled Pspto_3016 from *P. syringae* were obtained by hanging drop vapor diffusion at 18 °C from drops containing a 1:1 mixture of protein solution (7.5 mg/mL) and well-precipitant solution. The protein solutions contained 100 mM NaCl, 5 mM DTT, 0.02% NaN_3, 10 mM Tris-HCl (pH 7.5) and the reservoir solution contained 100 mM monosodium citrate (pH 4.2), 20% PEG 8000 and 60 mM $NaNO_3$. The crystal was cryo-protected with 20% glycerol and flash-cooled in liquid nitrogen. Diffraction data sets were collected on a single crystal using the beam line X4A with a Quantum 4R detector at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Data were integrated and scaled with HKL2000 (51). The crystal structure of Pspto_3016 was solved by SAD using SHELX (60). The locations of three selenium sites per molecule were identified from 2.9 Å resolution data. Matthews coefficient calculations indicated four monomers per asymmetric unit. Initial phases obtained using SHELXD, were improved by density modification and solvent flattening (using a solvent content of 46%). After phase refinement, an initial model constructed with RESOLVE (63) was extended by ARP/wARP (55) to 2.5 Å and refined with CNS (10). Model building was performed using Coot (19). Several cycles of simulated annealing and minimization were carried out using the CNS program package (10). Atomic coordinates were deposited in the Protein Data Bank (PDB ID 3H9X).

4.2.2 Results and Discussion

The solution NMR structural ensemble of Pspto_3016 (PDB ID 2KFP) is shown in Figure 1A. The structure adopted a characteristic α/β sandwich fold similar to that of the two other proteins with known structures from Pfam PF04237, revealing the putative "double wing" DNA binding motif (40, 62). The secondary structure is composed of three α-helices (α1-3), a solvent exposed β-sheet consisting of six anti-parallel strands (β1-6), and several loops with a secondary structure arrangement in the order N-α1-β1-β2-β3-β4-α2-β5-β6-α3-C (Figure 1B). NMR structure quality statistics are given in Table 1.

The X-ray crystal structure of Pspto_3016 (PDB ID 3H9X) is shown aligned to the NMR structure in Figure 1C, and the corresponding structural refinement statistics are shown in Table 2. The NMR and crystal structure are very similar to one another, with backbone RMSD of 1.39 ± 0.14 Å for residues with ordered secondary structure (2-14, 19-20, 29-32, 41-47, 58-65, 70-77, 80-82, 92-95, and 101-116), between the ensemble of 20 solution NMR models and the X-ray model. The ^{1}H-^{15}N resonances for residues Tyr40, Lys88, and Leu96 had lower peak intensity than average compared to other amide ^{1}H-^{15}N cross peaks. Trp39 and Ala83-Asn87 amides, and Tyr40 side chain resonances could not be assigned. Trp39 and Tyr40 residues reside on the loop separating strands β2 and β3, and residues Ala83-Lys88 are located on the extended loop connecting strands β5 and β6. These loop regions are likely to have considerable flexibility in Pspto_3016 and are represented with sausage diagrams created according to C^{α} RMSD for the NMR ensemble and according to B-factor for the X-ray structure in Figures 4.3D and 4.3E, respectively. The sausage diagram of the NMR model (Figure 4.3D) reflects the increase in backbone RMSD for the β2-3 and β5-6 loops due to incomplete resonance assignments and therefore a lack of structural constraints in these regions of the structure. Analysis of the sausage diagram for the crystal structure in the region of these loops indicated that the B-factors for residues in the β2-3 loop were slightly higher than in structured regions and that the B-factors for residues in the β5-6 loop region in the crystal structure (Figure 4.3E) were not unusually high compared to structured regions in the protein. Analysis of the crystal packing indicated that Trp39 in the β2-3 loop was in close proximity (< 4 Å) from Glu69 in a symmetry related molecule, and that residues Tyr84-Met86 of the β5-6 loop are in close proximity (< 5 Å) to Pro68-Glu69 of a symmetry mate, indicating that crystal packing may limit structural disorder in both the β2-3 and β5-6 loops. Additionally, the position of helix α2 is slightly different in the NMR ensemble compared to the X-ray structure, again indicating a possible consequence of packing against the β5-6 loop, which would otherwise experience conformational variability.

Submission of the atomic coordinates for the NMR ensemble to the Dali (31) server identified two structures in the Protein Data Bank of orthologous bacterial proteins

from Pfam PF04237 as structurally similar. The crystal structure of 132 residue hypothetical protein DR2400 from *D. radiodurans* (PDB ID 2A1V) was the most similar with a Z-score of 9.9 (seq ID = 32% and BLAST *E*-value = 4 x 10^{-4} obtained from the Kegg (38) database) followed by the NMR structure of 118-residue hypothetical protein YjbR (PDB ID 2FKI; NESG ID ER226) (62) from *E. coli* with a Z-score = 9.1 (seq ID = 33% and BLAST *E*-value = 4 x 10^{-4}). The primary difference between these structures and Pspto_3016 is that DR2400 and YjbR lack the pronounced loop segment between strands β3 and β4 and contain a short additional α-helix motif at the C-terminus (Figure 1F). Pspto_3016, like YjbR and DR2400, contains the highly conserved Asn-Lys-X-His-Trp motif (62) found on the loop connecting strands β5-β6. This sequence motif is characteristic of many members of Pfam PF04237. Further analysis of the Dali results identified the crystal structure of the 107-residue MotCF domain (PDB ID 1KAF) (40) of transcription factor MotA from bacteriophage T4, which has a lower Z-score of 4.9 but still exhibits many structural similarities despite a low sequence identity of 10%. The X-ray crystal structure of MotCF (PDB ID 1KAF) revealed a novel DNA-binding α/β motif, composed of three α-helices and a six-stranded β-sheet (40), which together with the activation domain MotNF, function as the transcription factor MotA (20, 21, 41). In MotCF, the putative DNA binding region of the double wing motif is composed of many electropositive and aromatic residues on the solvent exposed β-sheet, flanked by the two extended loop regions connecting strands β1-β2 and β5-β6, forming a binding pocket that resembles the shape of a saddle (40). MotCF displays a similar fold to the aforementioned α/β sandwich present in Pspto_3016. The two extended loops separating strands β1-β2, and between β5-β6, are structurally conserved features in Pspto_3016 as well as in the structures for proteins DR2400 and YjbR. Analysis of the electrostatic surface potential of Pspto_3016 with APBS (5) reveals a highly positively charged surface corresponding to the loops and β-sheet of the saddle (Figure 2A), including the surface surrounding the strongly-conserved Asn-Lys-X-His-Trp motif (Figure 2B). In MotCF, the β5-β6 loop is predicted to interact with the DNA major groove (40). The sequence Asn-Gly-Asn-Val-Tyr of MotCF, structurally aligns with the Asn-Lys-X-His-

Trp motif on the β5-β6 loop of Pspto_3016, and includes Tyr191 on the β6 strand, a predicted DNA-binding residue (40). In MotCF, additional residues, Tyr134 and Arg135 on the loop between strands β1-β2 and Lys166 on the C-terminal end of strand β4 (Figure 2C) are also predicted to interact with DNA (40)]. These structurally align with Trp22, Lys24 and Lys65 in Pspto_3016 (Figure 2D).

A CLANS (22) cluster analysis of the 1,845 Uniprot sequences from Pfam PF04237 revealed two major groupings of closely related sequences in the family (Figure 3). Using a CLANS *p*-value cutoff of 1×10^{-10}, Pspto_3016 was found in a distinct sub-family of 374 sequences, whereas YjbR was found in a separate sub-family with 1,073 other sequences (Figure 3), including the 33% identical protein DR2400. The sequences of each sub-family were extracted to generate multiple alignment files using ClustalW (64)], and submitted to the ConSurf server (3)] to map conserved surface residues from each sub-family onto the corresponding structures of Pspto_3016 and YjbR (Figures 2E and 2F). The location of the conserved surface patches for each sub-family was similarly located, surrounding the surface of the extended β5-β6 loop, presumably the site of DNA major groove interactions. Uniquely conserved in the Pspto_3016-containing sub-family was an insertion of residues 50-59 corresponding to the loop between strands β3-β4 (Alignment in Supplementary Figure S3). Additionally present was a Lys-Trp motif (Lys38 and Trp39) in the short loop connecting strands β2-β3 and an Ala-Tyr motif (Ala83 and Tyr84) of the β5-β6 loop (Figure 2B and 2E), which was not conserved in the YjbR sub-family. However, YjbR residues His21, Arg61, and His72 were conserved surface residues among its corresponding sub-family (Figure 2F). The conserved residues in each sub-family may be important for sequence specificity to the respective DNA substrate for the proteins from Pfam PF04237.

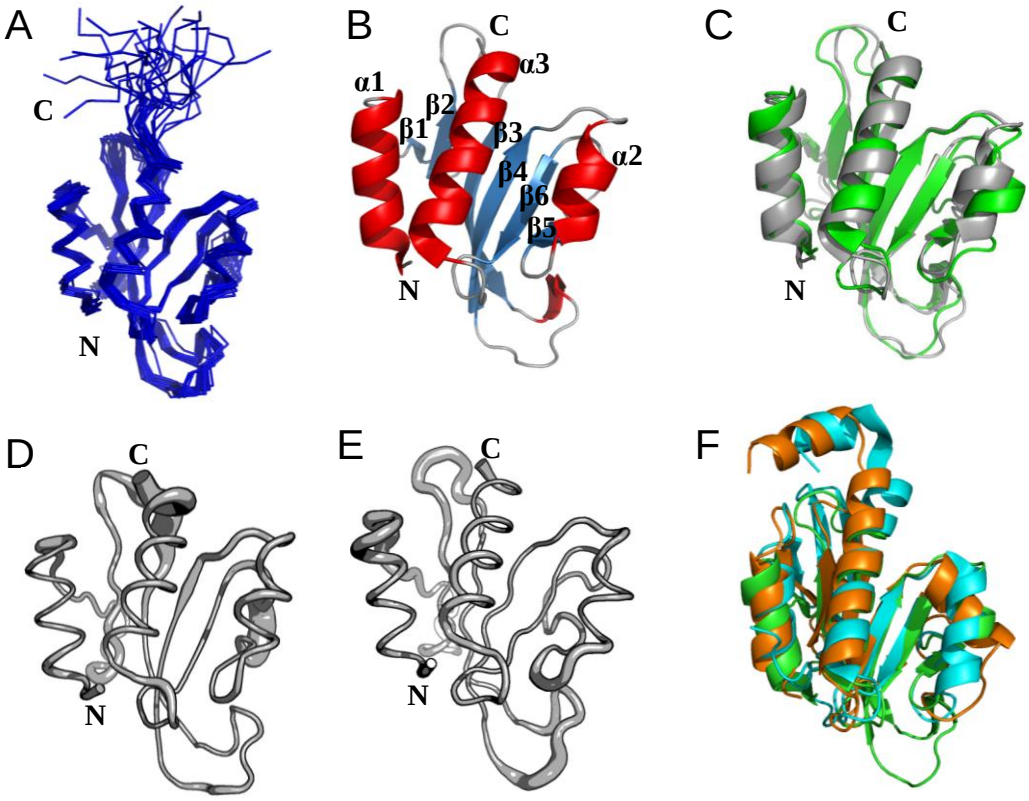

Figure 4.3 Three-dimensional structure of the Pspto_3016 protein from *P. syringae* (PDB IDs 2KFP and 3H9X). A) Backbone trace of the ensemble of 20 solution NMR conformers for residues 2-125 including the His_6 affinity tag. B) Ribbon diagram of the NMR structure for the conformer with the lowest energy from CNS refinement. Secondary structure elements are denoted α1-3 and β1-6 for respective features. C) Alignment of the X-ray structure (green) with the lowest energy NMR conformer (gray). D) Sausage diagram of the NMR ensemble according to the RMSD in backbone C^{α} position represented as the thickness of the trace. E) Sausage diagram, of the X-ray structure where the thickness of the trace is reflected by increasing backbone B-factor (PyMOL script from PDBe). F) Alignment of the Pspto_3016 X-ray structure (green) with the X-ray structure of DR2400 (PDB ID 2A1V) from *D. radiodurans* (cyan) and NMR structure of YjbR (PDB ID 2FKI) from *E. coli* (orange). Unless otherwise noted, residues 2-117 are shown for Pspto_3016, residues 3-130 for DR2400, and residues 2-118 for YjbR. All structures were rendered and alignments performed with PyMOL (58).

Completeness of resonance assignments[a]	
Backbone (%)	93.3
Side chain (%)	91.8
Aromatic (%)	83.0
Stereospecific methyl (%)	100.0
Conformationally-restricting constraints[b]	
Distance constraints	
Total	1822
Intra-residue [$i = j$]	418
Sequential [$\| i - j \| = 1$]	481
Medium range [$1 < \| i - j \| < 5$]	370
Long range [$\| i - j \| \geq 5$]	553
Dihedral angle constraints	172
Hydrogen bond constraints	80
Number of constraints per residue	17.0
Number of long range constraints per residue	4.9
Residual constraint violations[b]	
Average number of distance violations per structure	
0.1 - 0.2 Å	11.9
0.2 - 0.5 Å	3.5
> 0.5 Å	0.0
Average number of dihedral angle violations per structure	
1 - 10°	6.3
> 10°	0.0
RMSD from average coordinates (Å)[b,d]	
Backbone atoms	0.7
Heavy atoms	1.1
MolProbity Ramachandran statistics[b,d]	
Most favored regions (%)	98.5

Allowed regions (%)			1.2
Disallowed regions (%)			0.3
Global quality scores (Raw/ *Z*-score)[b]			
Verify3D	0.4	/	-1.1
ProsaII	0.6	/	-0.1
Procheck G-factor (phi-psi)[d]	0.0	/	0.3
Procheck G-factor (all)[d]	0.0	/	-0.2
MolProbity Clashscore	24.5	/	-2.7
RPF Scores[c]			
Recall / Precision	0.90		0.80
F-measure / DP-score	0.85		0.74

[a] Refers to chemical shifts for residues 2-117.

[b] Calculated for the ensemble of 20 structures using PSVS version 1.4 (6). Average distance violations were calculated using the sum over r^{-6}.

[c] RPF scores (33) calculated for the ensemble of 20 structures reflecting the goodness-of-fit to the NOESY data and resonance assignments.

[d] Ordered residue ranges: 2-36,40-53, 57-83, 90-99, 101-116, with the sum of φ and ψ order parameters > 1.8.

Table 4.1 Summary of NMR and structural statistics for Pspto_3016 (PDB ID 2KFP)

Data collection			
Space group	P1		
X-ray wavelength (Å)	0.979		
Unit-cell parameters	a =44.302 Å, b = 48.546 Å, c = 68.047 Å		
	α = 87.07°, β = 92.40°, γ = 93.57°		
Resolution (Å)	30-2.5 (2.5-2.59)		
Temperature (K)	100		
Completeness (%)	96.9 (93.5)		
Redundancy	2.0 (1.8)		
I/σ (I)	12.0 (2.3)		
R_{merge}	0.044 (0.255)		
Refinement			
R_{cryst}/R_{free}* =	0.210/0.245		
No. of subunits in ASU	4		
No. of residues	467		
No. of water molecules	38		
RMSD in bond lengths (Å)	0.010		
RMSD in bond angles (°)	1.30		
Ramachandran statistics			
Most favored regions (%)	95.7		
Allowed regions (%)	3.5		
Disallowed regions (%)	0.9		
Global quality scores (Raw / *Z*-score)			
Verify3D	0.6	/	1.4
ProsaII	1.0	/	1.4
Procheck G-factor (phi-psi)	-0.3	/	-0.8

Procheck G-factor (all)	-0.4	/	-2.4
MolProbity Clashscore	30.1	/	-3.6

Values in parantheses are representative of the highest resolution bin.

R_{merge} = Σ_{khl} Σ_i | $I_i(hkl)$ - <$I(hkl)$> |/ Σ_{khl} $\Sigma_i I_i$ (hkl), where $I(hkl)$ is the intensity of reflection hkl, Σ_{khl} is the sum over all reflections and Σ_i is the sum over I measurements of reflection hkl.

R^* = Σ_{khl} | | F_{obs} | - | F_{calc} | |/ Σ_{khl} $I_i(hkl)$ | F_{obs} |; R_{free}^* is calculated using 10% of the reflection data omitted from refinement of resonance assignments and R_{cryst} is calculated for the remaining reflections.

Table 4.2 Summary of crystals, X-ray diffraction data collection, and refinement statistics for Pspto_3016 (PDB ID 3H9X)

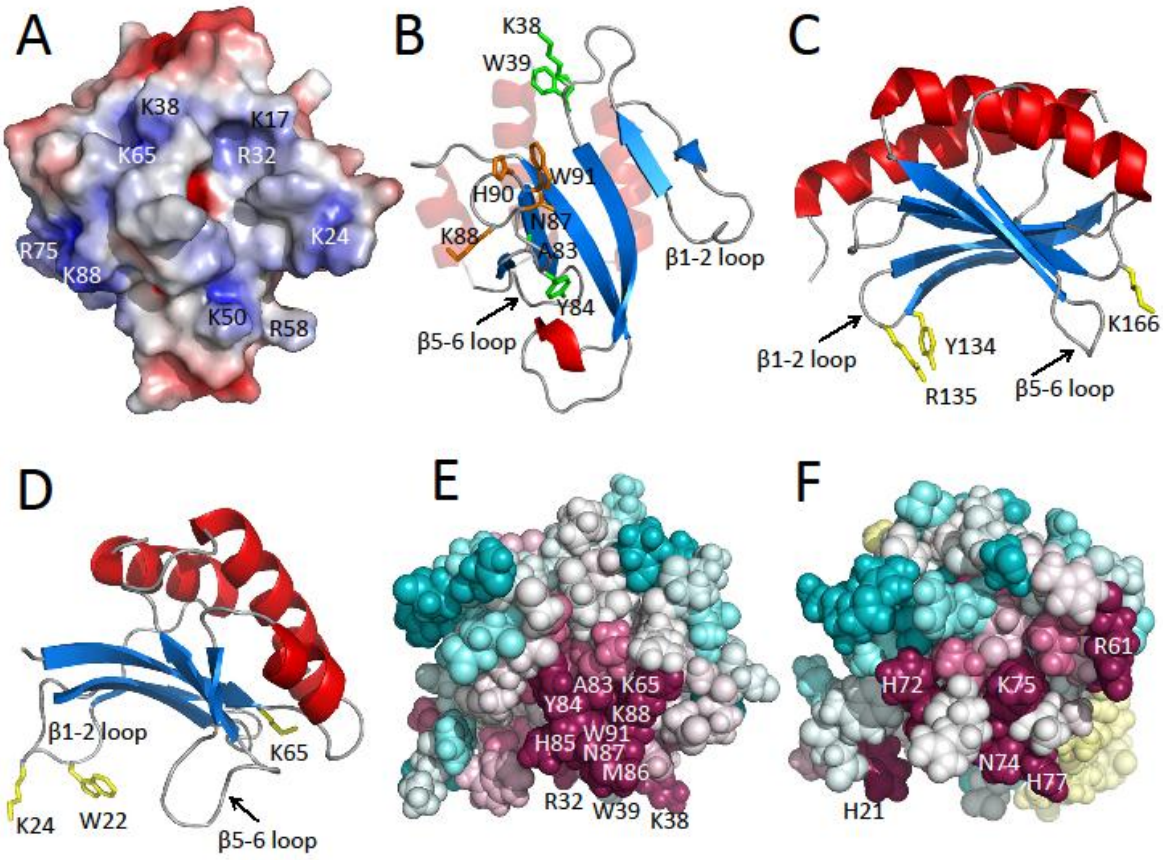

Figure 4.4 A) Electrostatic surface potential map of the lowest energy NMR conformer of Pspto_3016 generated with APBS (5) where negatively charged regions are represented in red, neutral regions in white, and positively charged regions in blue (± 5 kT/e). Locations of alkaline surface residues are indicated. B) Ribbon diagram with highlighted residue side chains represented as sticks. Residues Asn87, Lys88, His90, and Trp91 of the highly conserved Asn-Lys-X-His-Trp motif, are shown in orange, residues Lys38, Trp39, Ala83, and Tyr84 are shown in green. Pspto_3016 is shown in the same orientation as panel A. C) Ribbon diagram of MotCF (PDB ID 1KAF; residues 105-211 displayed) shown in the saddle orientation with highlighted putative DNA binding

residues Tyr134, Arg135, and Lys166 in yellow. D) Ribbon diagram of Pspto_3016 in the saddle orientation with residues Trp22, Lys24, and Lys65 highlighted in yellow. E) Pspto_3016 ConSurf (3) representation shown with spheres, using the aligned Pfam PF04237 sequences clustered from the CLANS (22)] sub-family. Conserved surface residues Arg32, Lys38, Trp39, Lys65, Ala83, Tyr84, His85, Met86, Asn87, Lys88, and Trp91 are indicated. E) ConSurf representation of YjbR from *Escherichia coli* using the corresponding CLANS sub-family sequences. YjbR was aligned to Pspto_3016 in the same orientation as panels D and E. Residues N74, K75, and H77 of the Asn-Lys-X-His-Trp motif are indicated on the surface and Trp78 is buried. Additional conserved residues His21, Arg61, and His72 are indicated.

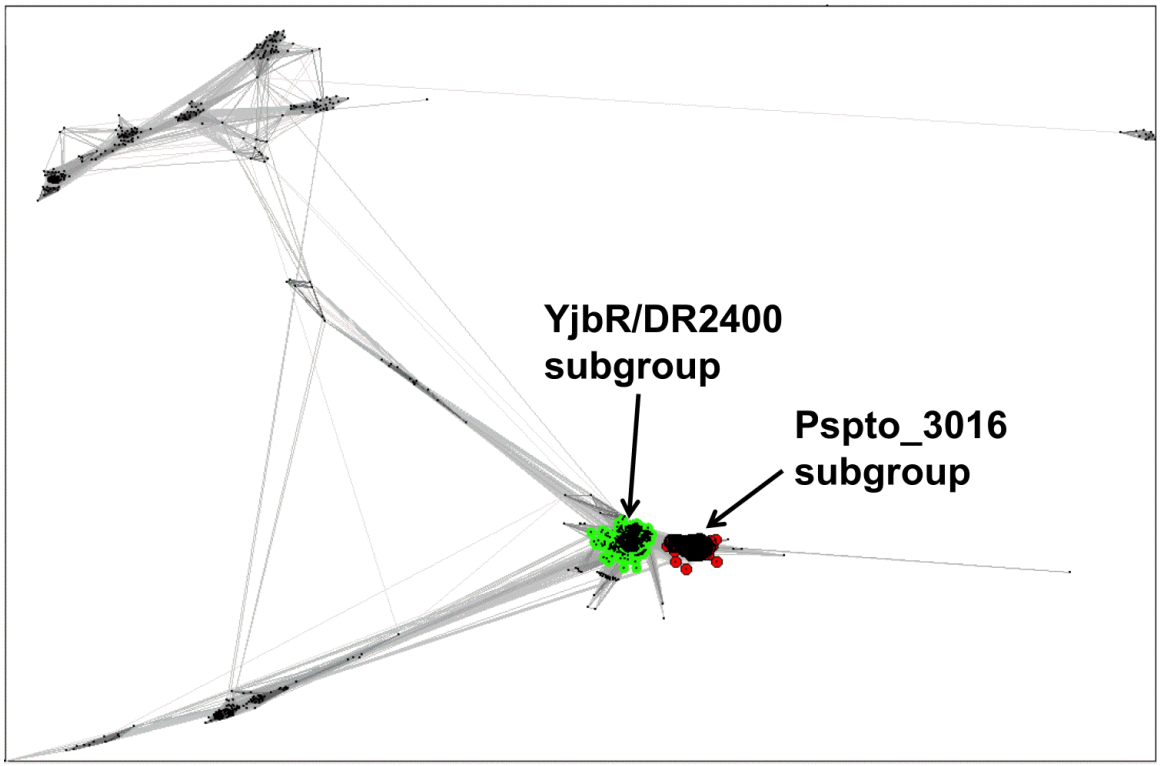

Figure 4.5 Cluster map of the Pfam PF04237 protein family using CLANS (22). Clusters were determined using a BLAST *p*-value cutoff of 1 x 10^{-10} to run the CLANS all-to-all pairwise comparison of the 1,845 sequences for PF04237 from the Uniprot database. The cluster containing Pspto_3016 is indicated on the right, and the cluster containing YjbR and DR2400 is on the left.

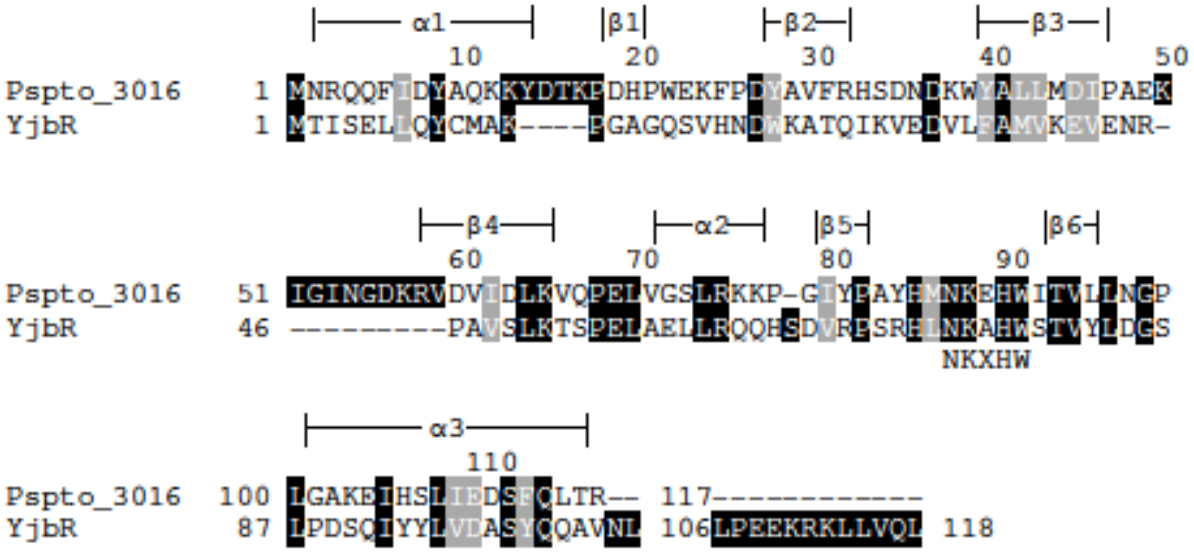

Figure 4.6 Sequence alignment of Pspto_3016 from *Pseudomonas syringae* with YjbR from *Escherichia coli*. Alignment was made using BOXSHADE v. 3.21. Secondary structure elements for Pspto_3016 are labeled above the residue sequences. The conserved Asn-Lys-X-His-Trp motif from Pfam PF04237 is indicated.

In summary, we have presented the solution NMR and X-ray crystal structures of Pspto_3016 from *P. syringae*, a member of the uncharacterized protein family PF04237 (DUF419). Sequence homology and structural similarities led to the identification of the MotCF domain of transcription factor MotA from T4 bacteriophage as a potential functional homolog and the putative function as a DNA-binding protein. In addition, other double wing proteins from Pfam family PF04237 with known three-dimensional structures: YjbR from *E. coli*, and DR4200 from *D. radiodurans*, are structurally similar and a high degree of conservation for several exposed electropositive and aromatic residues in the potential DNA-binding regions, including the highly conserved Asn-Lys-X-His-Trp motif (62). These observations as well as the overall structural similarity with the known DNA-binding protein MotCF, strongly suggest that Pspto_3016 is a sequence-specific DNA binding protein, and that this molecular function is common to members of the PF04237 domain family. However, since MotCF is the putative DNA-binding domain of the larger MotA protein that utilizes the activation domain MotNF for associating with other protein complexes, the independent gene product Pspto_3016 may require other yet unknown proteins in order to form complexes with DNA to carry out its physiological function.

4.3 Introduction to NESG target ID: NsR244

Here we present the solution NMR structure of the small (86-residue), acidic (pI 4.37) Asl3597 protein from *Nostoc* sp. PCC7120 (UniProtKB ID, Q8YR53_ANASP; NESG ID, NsR244; KEGG ID, ana:asl3597), a member of the (Pfam) PF12095 (DUF3571) protein domain family. Asl3597 was selected as a target by the Northeast Structural Genomics Consortium (NESG) as part of the National Institutes of Health Protein Structure Initiative-2, a massive international campaign to determine three-dimensional structures of broadly conserved protein domain families by NMR spectroscopy and X-ray crystallography. The PF12095 Pfam consists of proteins that are typically 85 to 95 amino acid residues in length found in heterocystous and non-heterocystous cyanobacteria, and upwards of 160 residues in plants due to the existence of a putative N-terminal 68-residue targeting sequence. The function of these proteins is still mostly unexplored with the exception of the 156-residue chlororespiratory reduction 7 (CRR7) protein from *Arabidopsis thaliana* (UniProtKB ID, Q9FL87_ARATH; KEGG ID, ath:AT5G39210, 33% identity). CRR7 is a soluble, chloroplast stromal protein required for assembly of the hydrophilic subcomplex of the chloroplast NAD(P)H dehydrogenase (NDH) complex together with CRR6 (37, 52). The NDH complex in higher plants is located in the thylakoids where it is involved in chlororespiration and cyclic electron transport (CET) by interacting with photosystem-I (PSI) to form the NDH-PSI supercomplex as shown in *Arabidopsis* (53, 54). The NDH complex in cyanobacteria (NDH-1) is also responsible for CO_2 cellular uptake, in addition to respiration and PSI-mediated CET, and is believed to be the evolutionary ancestor of the chloroplast NDH (23, 50). The NDH-1 complex is composed of hydrophilic and membrane subcomplexes that together form an L-shape that is visible by electron microscopy (2). However, little is still known about the molecular mechanism of CRR7 and Asl3597 and the function they play in assembly or stabilization of the NDH complex in plants and cyanobacteria. The solution NMR structure of Asl3597 reported here will hopefully assist in the future biochemical understanding of this family of proteins from Pfam PF12095.

4.3.1 Materials and Methods

4.3.1.1 Expression, purification, and characterization of Asl3597

The gene for residues 1-86 of the Asl3597 protein from *Nostoc* sp. PCC7120 were cloned into a pET21-23C vector with an additional C-terminal His_6 tag (LEHHHHHH) (NESG plasmid ID, NsR244-21.9). The protein was expressed and purified using standard protocols of the NESG consortium in order to prepare [*U*-^{13}C, ^{15}N] and *U*-^{15}N, 5% biosynthetically directed ^{13}C (NC5) samples (1). In brief, [*U*-^{13}C, ^{15}N] and NC5 Asl3597 were expressed in *E. coli* strain BL21 (DE3) + pMGK, grown in MJ9 medium(35) containing $(U\text{-}^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. Cells were grown at 37 °C with vigorous shaking until mid-log phase growth then induced with 1 mM isopropyl-β-*D*-thiogalactopyranoside (IPTG) and incubated at 17 °C overnight for protein expression. The cells were harvested by centrifugation, lysed by sonication, and the soluble fraction recovered by high speed centrifugation according to standard NESG protocols (1). Asl3597 protein was first purified with a Ni-affinity column (HisTrap HP, 5 mL), followed by additional purification through a gel filtration column (HiLoad 26/60 Superdex 75) using an ÄKTAxpress™ system (GE Healthcare). The buffer used for Ni-affinity chromatography consisted of 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 40 mM imidazole, 1 mM TCEP and the protein was eluted in the same buffer with the addition of 500 mM imidazole concentration. The protein buffer was then exchanged for gel-filtration chromatography into 20 mM MES, pH 6.5, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM dithiothreitol, 0.02% NaN_3. Sample purity (>95%) and molecular mass (11.4 kDa for [*U*-^{13}C, ^{15}N] and 11.0 kDa for NC5 samples) were verified by SDS PAGE and MALDI-TOF MS, respectively. Samples were concentrated for NMR by centrifugation to a final concentration of 0.8 mM, and then the sample was made 10% v/v in D_2O. Asl3597 behaved as a monomeric protein by recording static-light scattering measurements in-line with gel filtration using a Shodex KW-802.5 analytical gel

filtration column (30 µL injection) coupled to a miniDAWN TREOS detector (Wyatt technologies) for light scattering at λ=690 nm, using a flow rate of 0.5 mL/min at 40°C (Supplementary Fig. S1). However, the isotropic overall rotational correlation time (τ_c) estimated from ^{15}N T_1 and $T_{1\rho}$ relaxation data were not consistent for a monomeric protein of 11 kDa. The relatively short $T_{1\rho}$ of 64 ms and relatively long τ_c of 12.1 ns are values more indicative of a 17 kDa protein, which would undergo nuclear relaxation more quickly (Supplementary Fig. S2). This would unlikely be the result of Asl3597 forming a dimer, but more likely the protein is undergoing chemical or conformational exchange since the light scattering data clearly suggests that Asl3597 is monomeric.

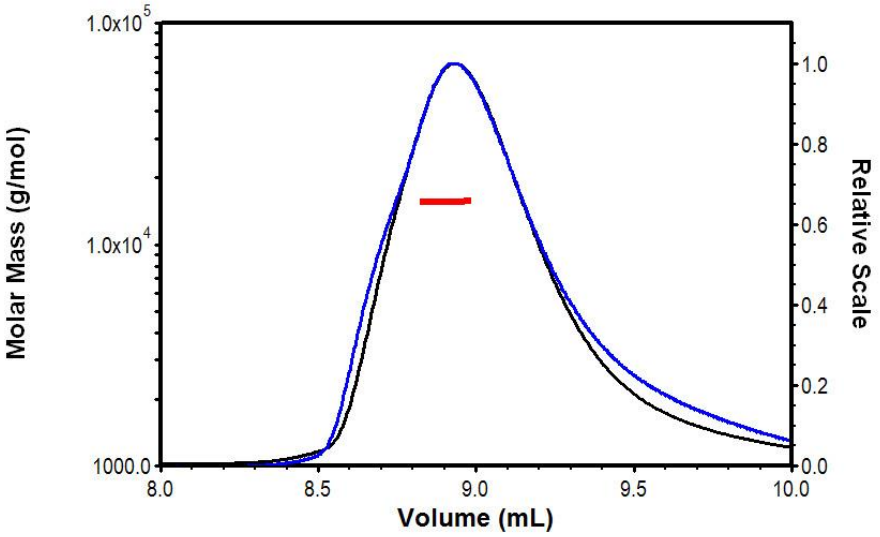

Figure 4.7 Static light scattering results for Asl3597. The NMR sample (30 µl) of [U-^{13}C^{15}N]-Asl3597 at 20 mM MES, pH 6.5, 200 mM NaCL, 5 mM $CaCL_2$, 10 mM dithiothreitol, 1X proteinase inhibitors, 0.02% NaN_3, 5% D_2O was injected onto an analytical gel-filtration column (KW-802.5, Shodex) with the effluent monitored by refractive index (black trace, Optilab rEX) and 90° static light-scattering (blue trace; miniDAWN TREOS, Wyatt Technology) detectors. The resulting experimental molecular weight of Asl3597 is 15.5 kDa (red), the expected MW including affinity tag is 11.6 kDa.

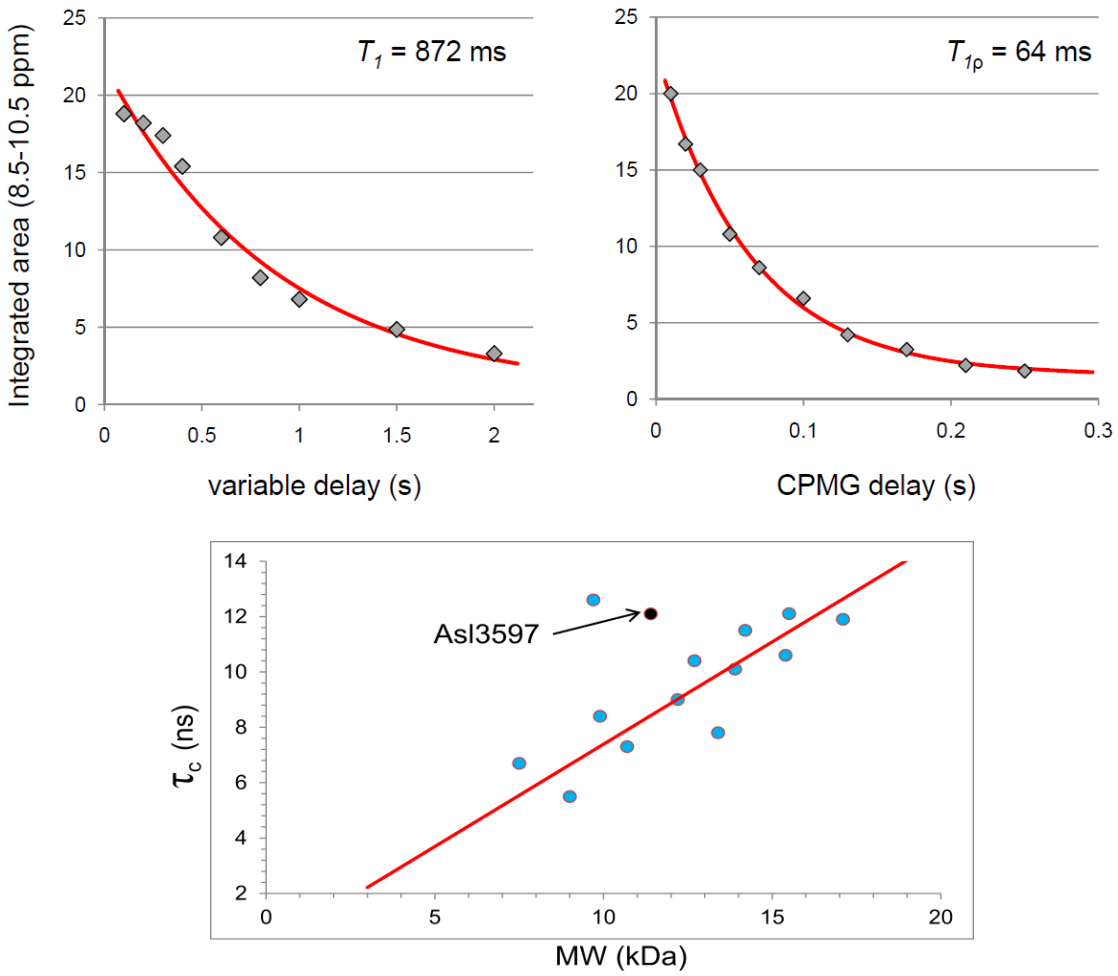

Figure 4.8 1D ^{15}N relaxation data for [*U*-^{13}C, ^{15}N]-Asl3597 acquired on a Varian Inova 600 MHz spectrometer at 20°C. (Top) Estimated T_1 and $T_{1\rho}$ relaxation times were obtained by integration from 8.5 to 10.5 ppm (using the intav macro) of the 1D ^{15}N-edited T_1 and $T_{1\rho}$ (CPMG) experiments and fit to an exponential decay (using the t1a or t2a macro, respectively). Longitudinal T_1 relaxation delays were 100, 200, 300, 400, 600, 800, 1000, 1500, and 2000 ms; transverse $T_{1\rho}$ relaxation CPMG delays were 10, 20, 30, 50, 70, 100, 130, 170, 210, and 250 ms; both experiments had 1.5 s recycle delays. (Bottom) Plot of average rotational correlation time, τ_c (ns), vs. protein molecular weight (kDa) of known monomeric proteins solved by solution NMR in the NESG consortium. The τ_c was approximated following derivation from the literature equation in Kay *et al.* (39) as described earlier.

4.3.1.2 NMR Spectroscopy

NMR data were collected at 20 °C on, [*U*-^{13}C, ^{15}N] and NC5 Asl3597 samples of approximately 270 µL in 5 mM Shigemi NMR tubes on 600 MHz Varian Inova and 850 MHz Bruker Avance III NMR spectrometers equipped with conventional 5-mM probes. The D_2O-exchanged samples were prepared by freezing [*U*-^{13}C, ^{15}N]-Asl3597, followed by overnight lyophilizing and resuspending into 99.9% D_2O (Acros Organics). Spectra were processed with NMRPipe (15) and visualized with Sparky 3.113 (27). Peptide backbone and side chain chemical shifts were determined from 2D ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC, and 3D HNCA, HNCO, HN(CO)CA, HNCACB, CBCA(CO)NH, HBHA(CO)NH, ^{15}N-edited NOESY-HSQC (τ_m = 70 ms), two ^{13}C-edited NOESY-HSQC (τ_m = 70 ms) experiments optimized for either aliphatic or aromatic carbons, H(C)CH-TOCSY, H(C)CH-COSY, H(CC)(CO)NH-TOCSY, (H)CC(CO)NH-TOCSY, and (H)CCH-TOCSY on [*U*-^{13}C, ^{15}N] samples. The 4D ^{13}C-^{13}C-HMQC-NOESY-HMQC (τ_m = 70 ms) experiment was collected on a D_2O-exchanged [*U*-^{13}C, ^{15}N] samples in order to assign additional NOEs. All 2D and 3D pulse sequences were from the Varian BioPack library and the 4D NOESY was from Lewis Kay (University of Toronto). The amide backbone $^{1}H^{N}$, $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$ and side chain $^{13}C^{\beta}$ and $^{1}H^{\beta}$ resonances were assigned after manual peak picking followed by data submission to the PINE server (NMRFAM) (4) for auto-assignment, while the side chain resonance assignments were completed manually. Stereospecific assignments of isopropyl side chain methyl groups of Leu and Val residues were performed using a constant time ^{13}C-HSQC experiments collected on the NC5 sample (49). The 2D ^{1}H-^{15}N HSQC spectrum of 0.8 mM [*U*-^{13}C, ^{15}N]-Asl3597 with assigned backbone and side chain resonances is shown in Figure 4.9. ^{15}N linewidths of amide backbone H-N peaks were measured in an unapodized ^{1}H-^{15}N HSQC spectrum using SPARKY, and were plotted for each residue (Figure 4.10). These linewidths were used to generate a sausage representation of Asl3597 shown in the bottom panel of Figure 4.10, where the broadest peaks are highlighted in red. The location of these amino acids clustered in and around the β-sheet face suggests these residues are undergoing chemical exchange. Therefore the observation of a longer than expected correlation time (τ_c) and

shorter than expected $T_{1\rho}$ relaxation time, is likely due to this chemical exchange phenomena. Chemical shifts, noesy peak lists, and raw FIDS were deposited in the BioMagResDB (BMRB accession number, 16652).

The solution NMR structure of Asl3597 was calculated with CYANA 2.1 (28, 30) using resonance assignments, NOESY peak lists with peak intensities in Sparky format, and dihedral angle constraints in order to automatically assign NOEs. Dihedral constraints for φ and ψ dihedral angles were derived using Talos+ (61). The 20 structures with lowest target function out of 100 were kept in each of 7 cycles. Resonance assignments and NOESY peak list cleanup was guided by NMR RPF (33) quality assessment scores used to assess the goodness-of-fit between the NOESY peak lists and the CYANA calculated structures. The NOE derived distance constraints from CYANA, the dihedral angle constraints derived from Talos+ (61), and hydrogen bond constraints derived manually, were converted to XPLOR / CNS format using PDBStat (6) with an increase of 10% to the upper bounds. These constraints were used to calculate the 100 structures using XPLOR-NIH-2.20 (59), followed by refinement in explicit water using CNS version 1.1 (42). Calculated structure statistics were made by submission to the PSVS server version 1.4 (Table 1) (6). These structures were used to identify consistent or flagrant NOE restraint violations for periodic manual assessment, leading to the eventual final NOE restraint list. The final ensemble of 20 models, sorted by the lowest energy, was submitted to the Protein Data Bank (PDB ID: 2KRX).

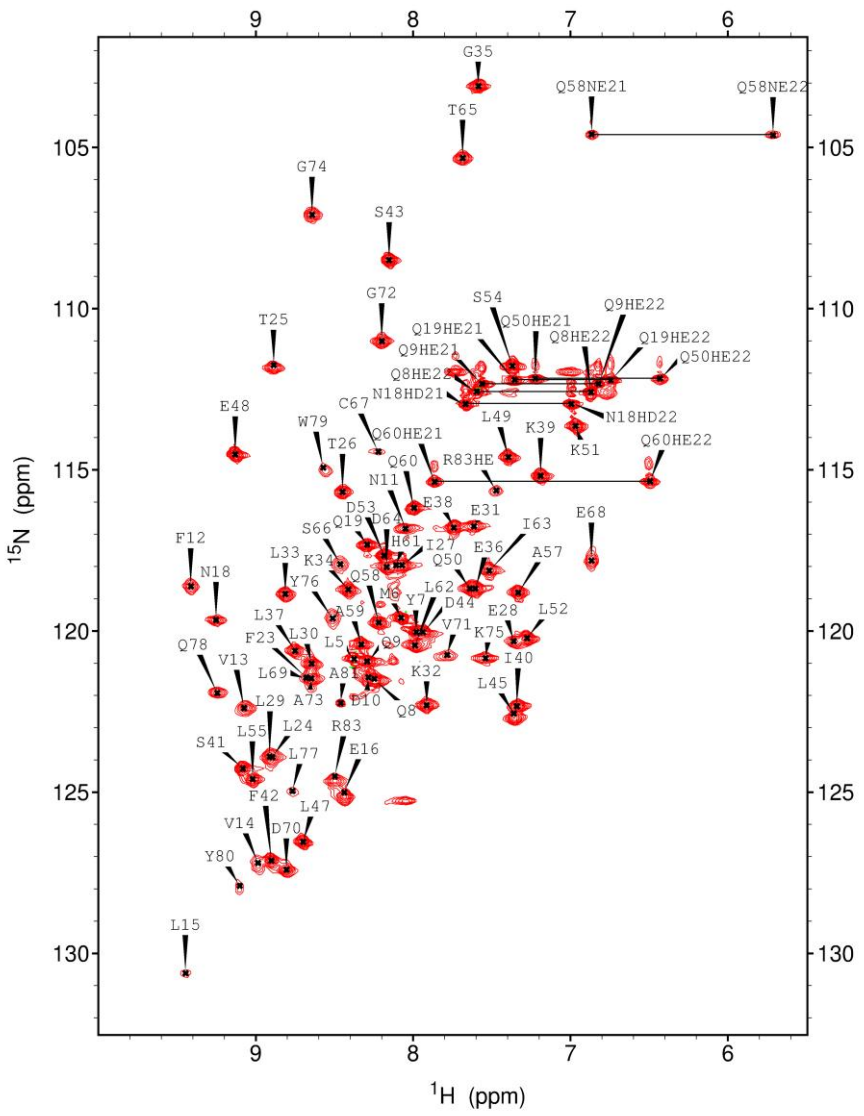

Figure 4.9 Two dimensional ^{1}H-^{15}N HSQC spectrum of 0.8 mM for [*U*-^{13}C, ^{15}N]-Asl3597 in a 10% v/v D_2O solution containing: 20 mM MES, pH 6.5, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM dithiothreitol, 0.02% NaN_3, acquired on a Varian Inova 600 MHz spectrometer at 20°C. Backbone resonance assignments are indicated above with their one-letter amino acid codes followed by their sequence number. Also labeled are the assigned side chain NH_2 resonances from Asn and Gln residues and the NH from Arg-83.

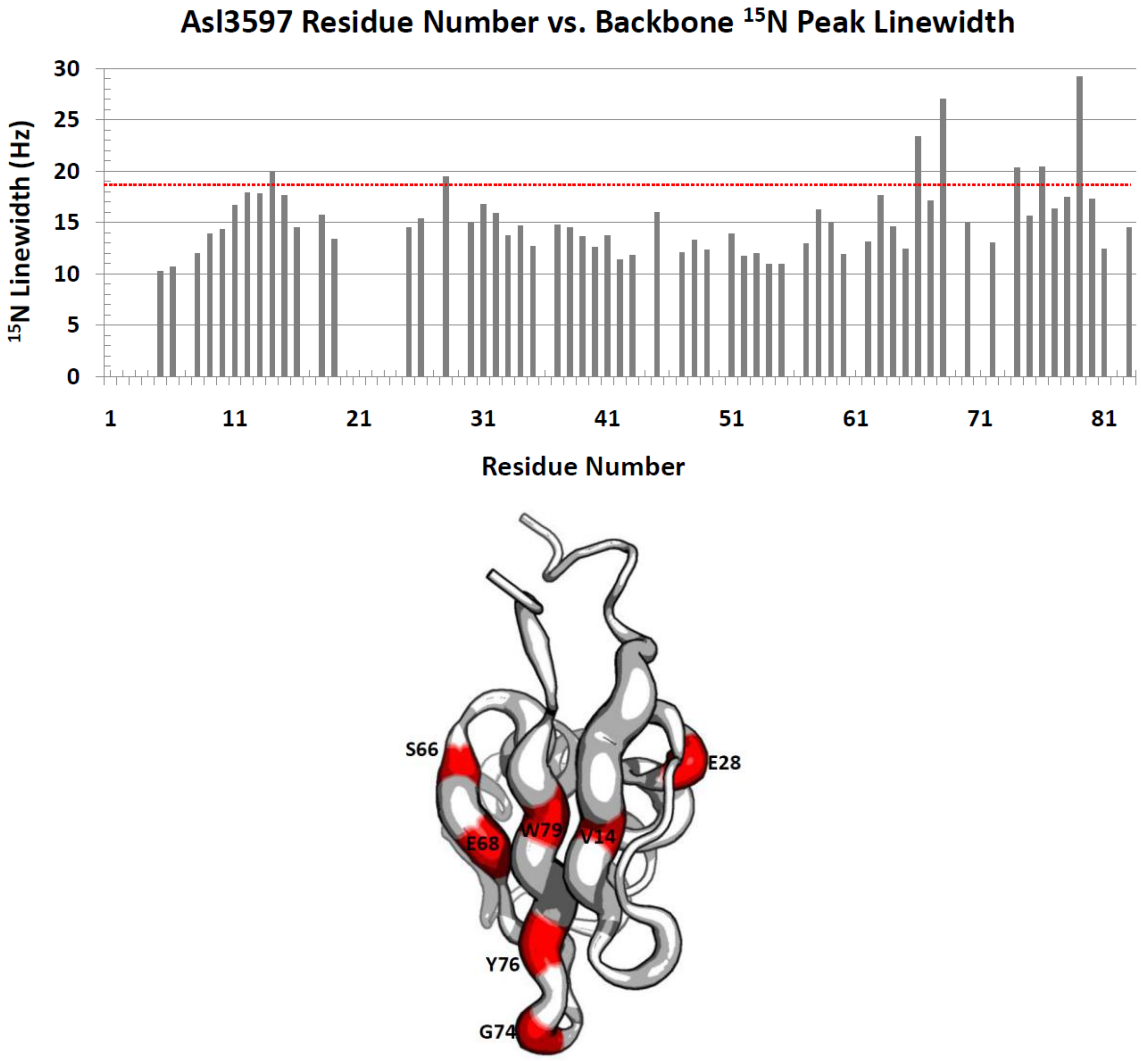

Figure 4.10 (Top) Bar graph of backbone ^{15}N peak linewidths (in Hz) of [*U*-^{13}C, ^{15}N]-Asl3597 for all unique amide H-N peaks in the ^{1}H-^{15}N HSQC spectrum (^{15}N dimension: sweep width = 2670.9 Hz, 256 real + imaginary points, $t_{1,max}$ = 51.2 ms, spectral resolution = 5.2 Hz after linear prediction and zero filling) acquired on a Bruker Avance III 850 MHz spectrometer at 20 °C. The dashed red line indicates the mean value plus the standard deviation (18.8 Hz) used for selecting residues with broad peaks exhibited by above average ^{15}N linewidths. (Bottom) Sausage representation of Asl3597 portraying backbone thickness as a function of ^{15}N linewidth. Highlighted in red are the residues above the 18.8 Hz threshold.

Completeness of resonance assignments[a]	
Backbone (%)	93.0
Side chain (%)	91.7
Aromatic (%)	87.1
Stereospecific methyl (%)	100.0
Conformationally-restricting constraints[b]	
Distance constraints	
Total	1570
Intraresidue ($i = j$)	349
Sequential [$(i - 1) = 1$]	340
Medium range [$1 < (i – j) < 5$]	423
Long range [$(i – j) \geq 5$]	458
Dihedral angle constraints	136
Hydrogen bond constraints	36
Number of constraints per residue	20.7
Number of long range constraints per residue	5.5
Residual constraint violations[b]	
Average number of distance violations per structure	
0.1 - 0.2 Å	2.7
0.2 - 0.5 Å	0.2
> 0.5 Å	0.0
Average number of dihedral angle violations per structure	
1-10°	5.1
>10°	0.0
RMSD from average coordinates (Å)[b,d]	
Backbone atoms	0.6
Heavy atoms	1.0

MolProbity Ramachandran statistics[b,d]			
	Most favored regions (%)		99.0
	Additionally allowed regions (%)		1.0
	Generously allowed regions (%)		0.0
	Disallowed regions (%)		0.0
Global quality scores (Raw/ *Z*-score)[b]			
	Verify3D	0.2	-4.7
	ProsaII	0.3	-1.3
	Procheck (phi-psi)[d]	0.0	0.4
	Procheck (all)[d]	0.0	-0.4
	Molprobity clash	23.0	-2.4
RPF Scores[c]			
	Recall / Precision	0.88	0.89
	F-measure / DP-score	0.89	0.75

[a] Refers to chemical shifts for residues 3-83.

[b] Calculated for the ensemble of 20 structures using PSVS version 1.4 (6). Average distance violations were calculated using the sum over r^{-6}.

[c] RPF scores (33) calculated for the ensemble of 20 structures reflecting the goodness-of-fit to the NOESY data and resonance assignments.

[d] Ordered residue ranges: 11-20, 23-65, 69-83, with the sum of φ and ψ order parameters > 1.8.

Table 4.3 Summary of NMR and structural statistics for Asl3597 (PDB ID 2KRX)

4.3.2 Results and Discussion

The solution NMR structure of Asl3597 is the first structure from Pfam domain PF12095, which remains a mostly uncharacterized family of proteins of unknown function. Asl3597 has a unique α/β sandwich fold consisting of four anti-parallel β-strands (β1, N11-E16; β2, Q19-T25; β3, L69-G72; β4, K75-A81) facing opposite a continuous three α-helix bundle (α1, T26-K39; α2, L47-K51; α3, L55-T65), and flanked by N- and C-terminal disordered tails with a secondary structure order of N-β1-β2-α1-α2-α3-β3-β4-C (Fig. 1A,B). Asl3597 has low sequence similarity with any other protein in the PDB (< 20%), and structural alignment using the Dali (31) server revealed that the closest structural similarity, which was still low, with small domains of RNA polymerase complex I and II from a variety of different source organisms (top result: RPO1N subunit of the archaeal 13-subunit DNA-directed RNA polymerase; PDB ID, 2WB1; Dali *Z*-score, 3.6; C^{α} RMSD, 4.6 Å) of which there also exists very low sequence identity (< 10%). Therefore it appears that Asl3597 exhibits an entirely novel protein fold.

Sequence analysis using a BLAST search provided by KEGG(38) yielded only proteins identified as either "hypothetical" or "uncharacterized" (*E*-values: 10^{1} - 10^{-35}) with the exception of the 156-residue CRR7 protein (*E*-value: 0.007, 33% identity) from *Arabidopsis thaliana*. Recent data indicated that CRR7, in conjunction with CRR6, is required for post-translational assembly of the hydrophilic subcomplex of the chloroplast NDH complex in *Arabidopsis*. CRR7 was found localized to the chloroplast stromal fraction, potentially facilitated by the N-terminal 68 residues that contain a predicted plastid targeting sequence (52). CRR7 is believed to play a role in biogenesis of the hydrophilic subcomplex based on the study of *crr7* null mutants that were discovered to have impaired NDH electron transport activity (37). The authors suggested that although CRR7 may be involved in the assembly of the hydrophilic subcomplex, it most likely is not a subunit of NDH (52). As expected from our BLAST search, the top results for Asl3597 were orthologous hypothetical proteins from other heterocyst forming and non-heterocystous cyanobacterial species all belonging to Pfam12095. Conservation of amino acid residues was assessed using BOXSHADE version 3.21 and the ConSurf

server (3), following generation of a multiple sequence alignment file using ClustalW (64). Sequence alignment across cyanobacterial proteins and CRR7 from *Arabidopsis* showed that stretches of homology exist across the entirety of the protein sequence including charged, aromatic, and nonpolar amino acids (Fig. 2). ConSurf analysis revealed two conserved surface patches, one at the disordered C-terminal tail from R83 to K86, and one that extends across β1 (L15), β2 (E21), and β4 (Q78-Y80) (Fig. 1C). Analysis of the electrostatic surface potential (5) revealed a mostly negatively-charged β-sheet face (Fig. 1D). Unfortunately due to the novel fold and uncharacterized nature of Asl3597, it is currently difficult to ascertain any other functional information beyond that derived from the solution NMR structure provided here.

In summary, we have determined the solution NMR structure of Asl3597 from *Nostoc* sp. PCC 7120, the first structure of the PF12095 (DUF3571) protein domain family, which remains functionally uncharacterized. Sequence analysis of Asl3597 led to identification of the ortholog CRR7 from *Arabidopsis thaliana,* which has recently been suggested to play a role in the biogenesis of the hydrophilic subcomplex of the NDH complex in the chloroplast stroma. Additional biochemical relevance for this unique structural motif of Asl3597, in particular from a photosynthetic bacterium, remains to be determined.

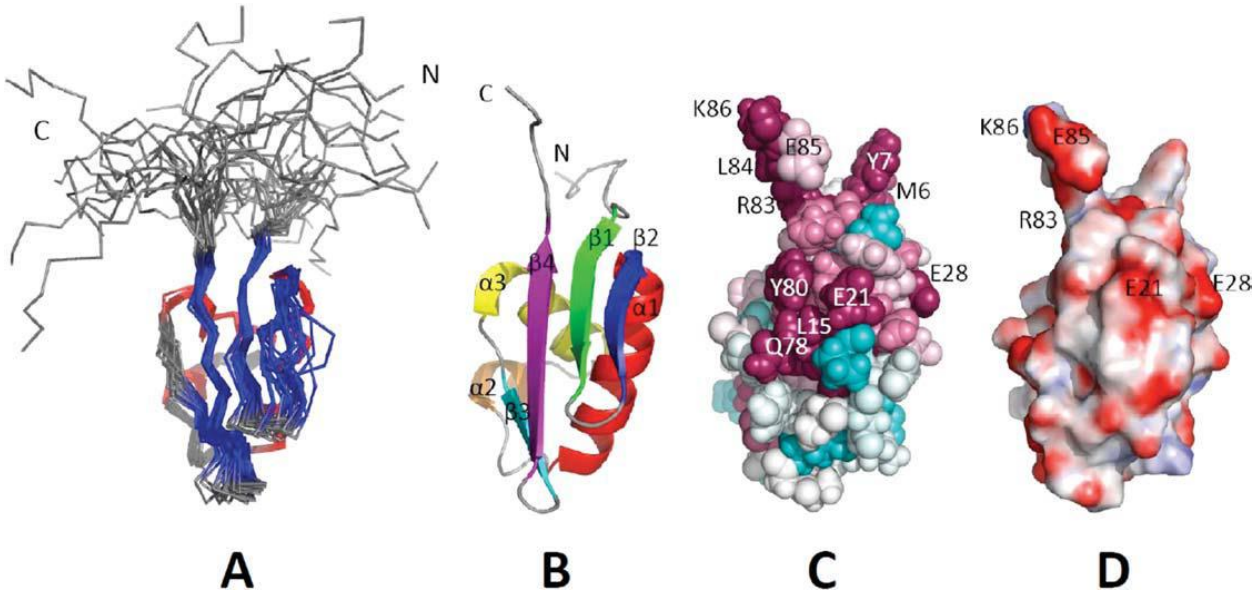

Figure 4.11 (A) Superposition of the final ensemble of 20 conformers from the solution NMR structure of Asl3597 (PDB ID, 2KRX). (B) Cartoon representation of Asl3597 showing β-strand order 3412. Secondary structure elements are labeled; α-helices are shown in red, β-strands are shown in blue, and loops are shown in gray. (C) ConSurf (3) image showing the conserved residues in Asl3597 from the top 10 BLAST results (KEGG (38) database) aligned with ClustalW (64). The color scheme reflects the degree of residue conservation over these members from protein domain family PF12095, and is depicted as completely conserved (magenta), highly conserved (dark and light pink), average (white), and variable (cyan). Selected highly conserved residues are labeled. (D) ABPS (5) image of the electrostatic surface potential of Asl3597. Positively charged are shown in blue, negatively charged shown in red and neutral shown in white (±20 kT/e). All structures were rendered using PyMOL molecular visualization software for residues 2-86.

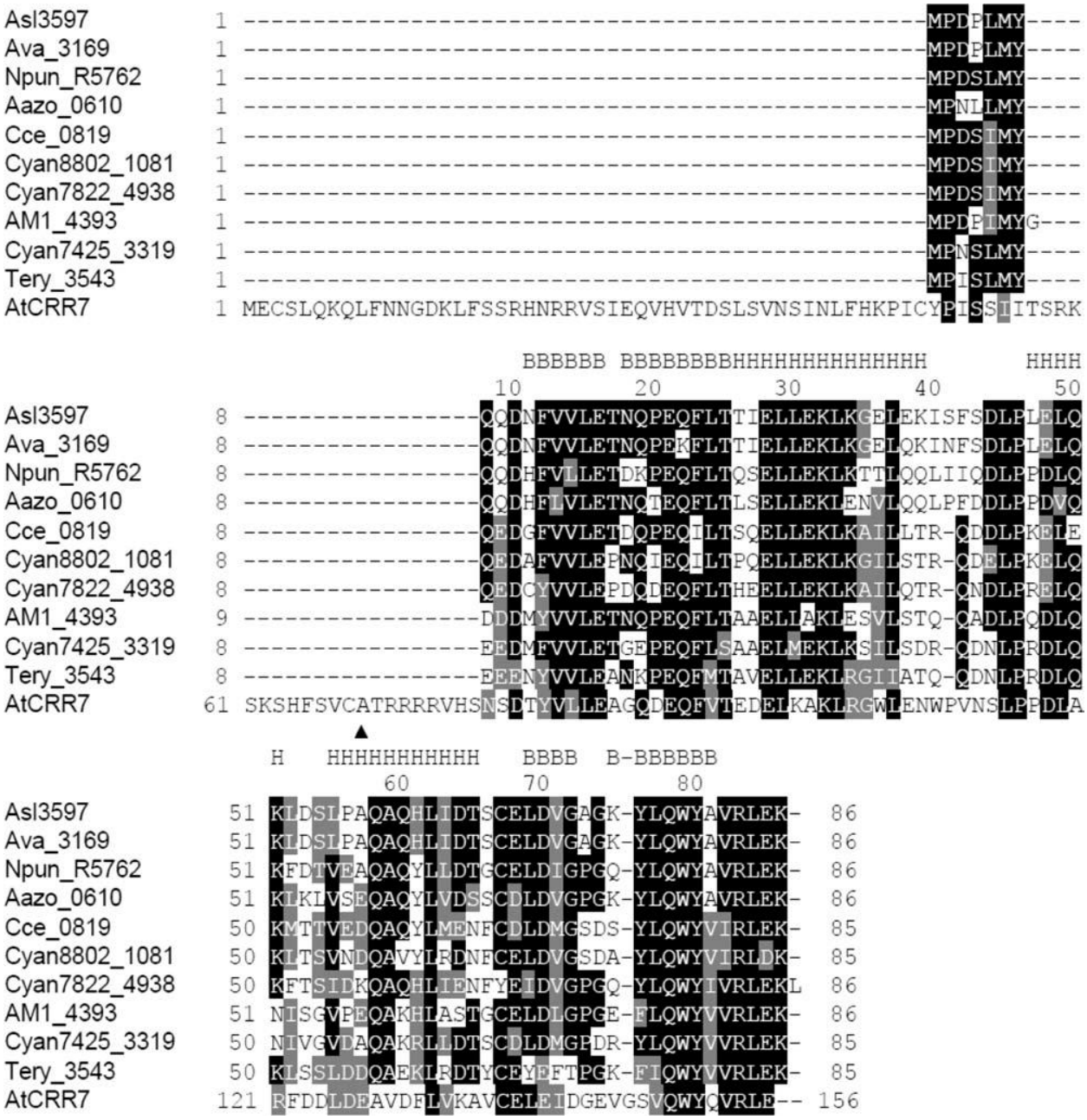

Figure 4.12 Alignment of Asl3597 from Nostoc sp. PCC7120 with the top 10 homologous BLAST (KEGG) results using BOXSHADE v. 3.21. CRR7 from Arabidopsis thaliana is also included at the bottom with an accompanying solid black triangle indicating the predicted cleavage site of the plastid targeting signal. Secondary structure elements from Asl3597 are indicated above the amino acid residue numbers in the sequence.

4.4 Introduction to NESG target ID: NsR143

Presented here is the solution NMR structure (PDB ID 2L09) of protein Asr4154 (NESG ID NsR143) from *Anabaena* sp. PCC 7120. Asr4154 (UniProtKB/TrEMBL ID Q8YPN9_NOSS1) is a member of the Pfam PF08369, which is annotated as the protochlorophyllide reductase (PCP_red) 57 kDa subunit family of bacterial and plant chloroplast proteins. Asr4154 adopts a three-helix bundle arrangement with disordered N- and C-termini. The solution NMR structures of Asr4154, and the 30% identical C-terminal protein fragment of BchB from *Chlorobium tepidum* (ctBchB-C) (NESG taget ID CtR69A; UniProtKB/TrEMBL ID BCHB_CHLTE, accession number Q9F715), were solved by the NESG simultaneously and together represent the first structures of Pfam PF08369.

4.4.1 Materials and Methods

Uniform ^{13}C, ^{15}N- labeling of Asr4154 was performed using standard protocols of the NESG consortium (1). In brief, the *asr4154* gene from *Anabaena sp.* PCC 7120 was cloned into the pET21-23C expression vector containing a non-native C-terminal affinity tag (LEHHHHHH), generating the construct NsR143-9-61-21.16. The vector DNA sequence was confirmed by sequencing analysis, transformed into BL21 (DE3) pMGK *E. coli* competent cells, and cells grown at 37 °C in MJ9 minimal medium. Cells were grown at 37 °C until OD_{600} of 0.6 then the temperature was shifted to 17 °C and then the culture was allowed to grow overnight following induction with isopropyl-β-*D*-thiogalactopyranoside (IPTG) before harvesting. Labeled Asr4154 protein was purified using Ni-NTA (Qiagen) affinity, followed by Superdex 75 (Amersham Biosciences) gel-filtration chromatography and verified by SDS-PAGE as a single strong band. *U*-^{15}N and 5% biosynthetically directed ^{13}C- (NC5) Asr4154 was also expressed and purified for assigning stereospecificity to isopropyl methyl groups of all Leu and Val residues (49). Final NC and NC5 NMR samples were concentrated to 0.8 mM in 95%/H_2O/5%/D_2O solution buffer containing 20 mM MES, pH 6.5, 100 mM NaCl, 10 mM DTT and 5 mM $CaCl_2$. Gel-filtration chromatography of Asr4154 with in-line static light scattering

confirmed the protein to be a monomer of approximately 7 kDa in molecular weight and verified by MALDI-TOF mass spectrometry (7.78 kDa for NC Asr4154).

All NMR experiments were recorded at 293 K on Varian INOVA 500 MHz, Varian INOVA 600 MHz, and Bruker AVANCE 850 MHz spectrometers. Proton chemical shifts were referenced to external DSS. 2D, 3D, and 4D data sets were processed with NMRPipe (15) and analyzed with Sparky 3.110 (27) software. Chemical shift assignments were made using the following double and triple resonance experiments recorded on the NC samples: 2D 15N-HSQC (Figure 4.13) and 13C-HSQC, 3D HNCA, HNCACB, HNCO, HN(CO)CA, CBCA(CO)NH, HBHA(CO)NH, H(C)CH-TOCSY, H(CC)(CO)NH –TOCSY, (H)CC(CO)NH-TOCSY, (H)CCH-TOCSY, ^{15}N-edited NOESY-HSQC, ^{13}C-edited NOESY-HSQC, and 4D ^{13}C-^{13}C-HMQC-NOESY-HMQC (τ_m = 70 ms for all NOESY experiments). Backbone amide ^{1}H, ^{15}N, $^{13}C^{\alpha}$ and side chain $^{13}C^{\beta}$ atoms were assigned after data submission to the PINE server (NMRFAM) (4) for auto-assignment whereas all others were assigned manually.

Dihedral angle constraints were computed by Talos+ (61) based on assigned chemical shifts. Preliminary structures were calculated with Autostructure 2.1.1 (34) and Cyana 2.1 (30) both of which were used to derive NOESY distance constraints from NOE cross peaks used in final structure calculations with XPLOR-NIH-2.20 (59). The final Xplor structures were further refined by restrained molecular dynamics in explicit water with CNS (version 1.1) (10) and sorted by the lowest energy. A final ensemble of the 20 lowest energy structures out of 150 calculated was deposited to the Protein Data Bank (PDB ID: 2L09). The RPF analysis program (33) was used to determine the global goodness-of-fit of the final ensemble with the final refined NOESY peak list data. Structures were analyzed by raw and Z-scores using the Verify3D, ProsaII, PROCHECK, and MolProbity statistics and global structure quality factors computed from PSVS version 1.4 (6). Final chemical shift assignments were deposited in the BioMagResDB (BMRB accession number 17035).

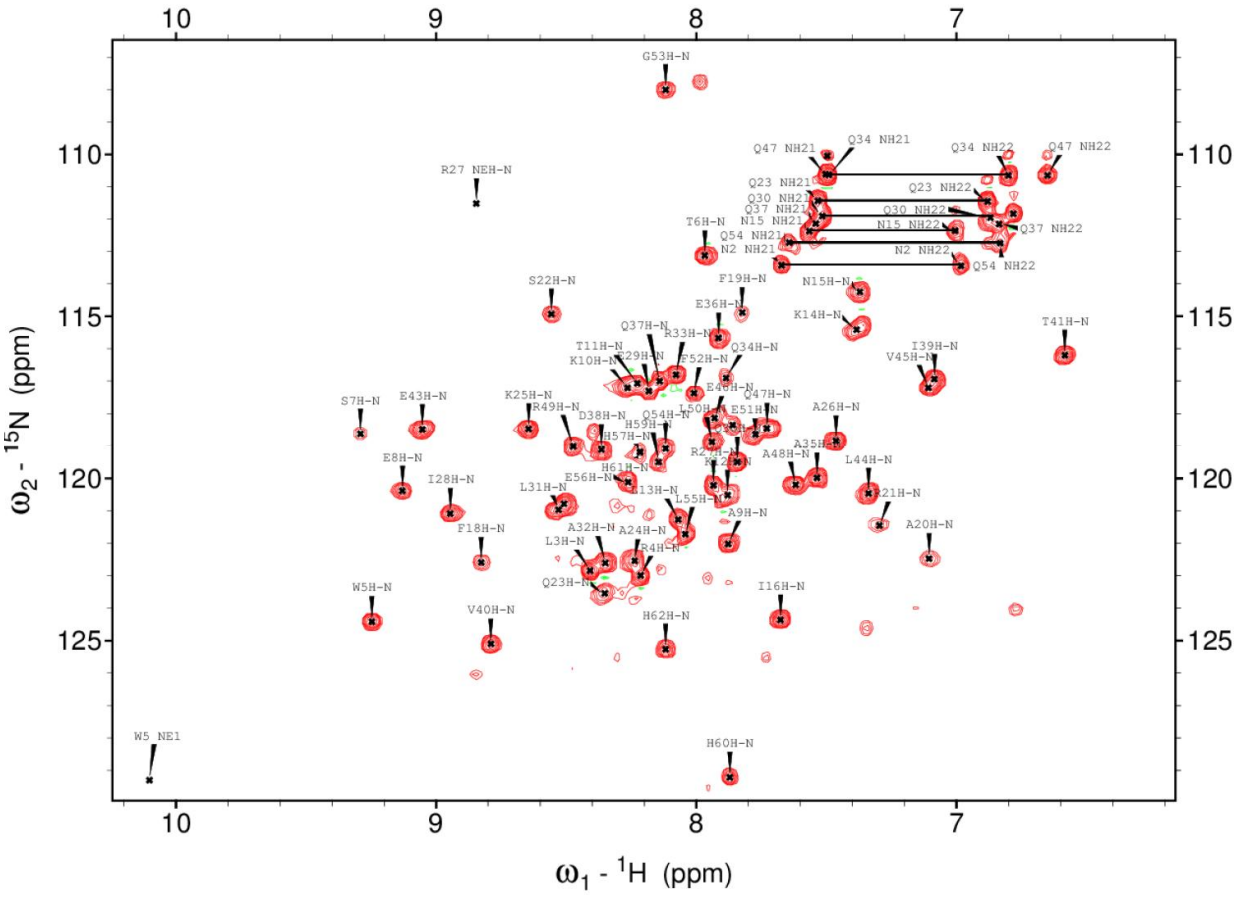

Figure 4.13 Two dimensional ^{1}H-^{15}N HSQC spectrum of 0.8 mM for [*U*-^{13}C, ^{15}N]-Asr4154 in a 10% v/v D_2O solution containing: 20 mM MES, pH 6.5, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM dithiothreitol, 0.02% NaN_3, acquired on a Bruker AVANCE III 850 MHz spectrometer at 20 °C. Backbone and side chain resonance assignments are indicated above with their corresponding one-letter amino acid codes followed by their sequence number and H-N pair type.

4.4.2 Results and Discussion

Presented here is the solution NMR structure of Asr4154 from *Anabaena* sp. strain PCC 7120 (PDB ID 2L09), and the corresponding structural statistics are presented in Table 4.4. The Asr4154 structure folds into a three-α-helix bundle arrangement including two short anti-parallel β-strands, in the secondary structure order of β1-α1- α2- β2- α3 (Figure 4.14). The atomic coordinates for Asr4154 were submitted to the Dali server (31) to scan for structurally similar proteins, at which point, a 54-residue C-terminal fragment (residues 484-537) of the much larger full-length BchB protein (1-537 residues) from *Chlorobium tepidum* was identified with a Dali *Z*-score of 7.1. The solution NMR structure of the BchB C-terminal fragment (ctBchB-C) (PDB ID 2KRU) was also solved by the NESG consortium (NESG target ID CtR69A) and released by the PDB six months apart from the Asr4154 structure in 2010. Together, the solution NMR structures of Asr4154 and ctBchB-C are the only representative structures from Pfam PF08369. The two proteins share 30% sequence identity but are very structurally similar, indicated by a backbone RMSD of 1.5 Å between the two lowest energy models (Asr4154 residues 2-54; ctBchB-C residues 484-537). The structural alignment for the models of Asr4154 and ctBchB-C is shown in Figure 4.14, panel D.

Completeness of resonance assignments[a]	
Backbone (%)	99.2
Side chain (%)	97.0
Stereospecific methyl (%)	100.0
Conformationally-restricting constraints[b]	
Distance constraints	
Total	1237
Intra-residue [$i = j$]	276
Sequential [$\| i - j \| = 1$]	264
Medium range [$1 < \| i - j \| < 5$]	333
Long range [$\| i - j \| \geq 5$]	364
Dihedral angle constraints	86
Hydrogen bond constraints	56
Number of constraints per residue	25.1
Number of long range constraints per residue	6.6
Residual constraint violations[b]	
Average number of distance violations per structure	
0.1 - 0.2 Å	1.2
0.2 – 0.5 Å	0.0
> 0.5 Å	0.0
Average number of dihedral angle violations per structure	
1 - 10°	0.5
> 10°	0.0
RMSD from average coordinates (Å) [b,d]	
Backbone atoms	0.3
Heavy atoms	0.9
MolProbity Ramachandran statistics [b,d]	

	Most favored regions (%)			99.9
	Allowed regions (%)			0.1
	Disallowed regions (%)			0.0
Global quality scores (Raw/ *Z*-score) [b]				
	Verify3D	0.3	/	-2.7
	ProsaII	0.8	/	0.7
	Procheck G-factor (phi-psi)[d]	0.3	/	1.6
	Procheck G-factor (all)[d]	0.2	/	1.4
	MolProbity Clashscore	20.6	/	-2.0
RPF Scores[c]				
	Recall / Precision	0.89		0.96
	F-measure / DP-score	0.92		0.79

[a] Refers to chemical shifts for residues 3-54.
[b] Calculated for the ensemble of 20 structures using PSVS version 1.4 (6). Average distance violations were calculated using the sum over r^{-6}.
[c] RPF scores (33) calculated for the ensemble of 20 structures reflecting the goodness-of-fit to the NOESY data and resonance assignments.
[d] Ordered residue range: 3-54 with the sum of φ and ψ order parameters > 1.8.

Table 4.4 Summary of solution NMR structure statistics for Asr4154 (PDB ID 2L09)

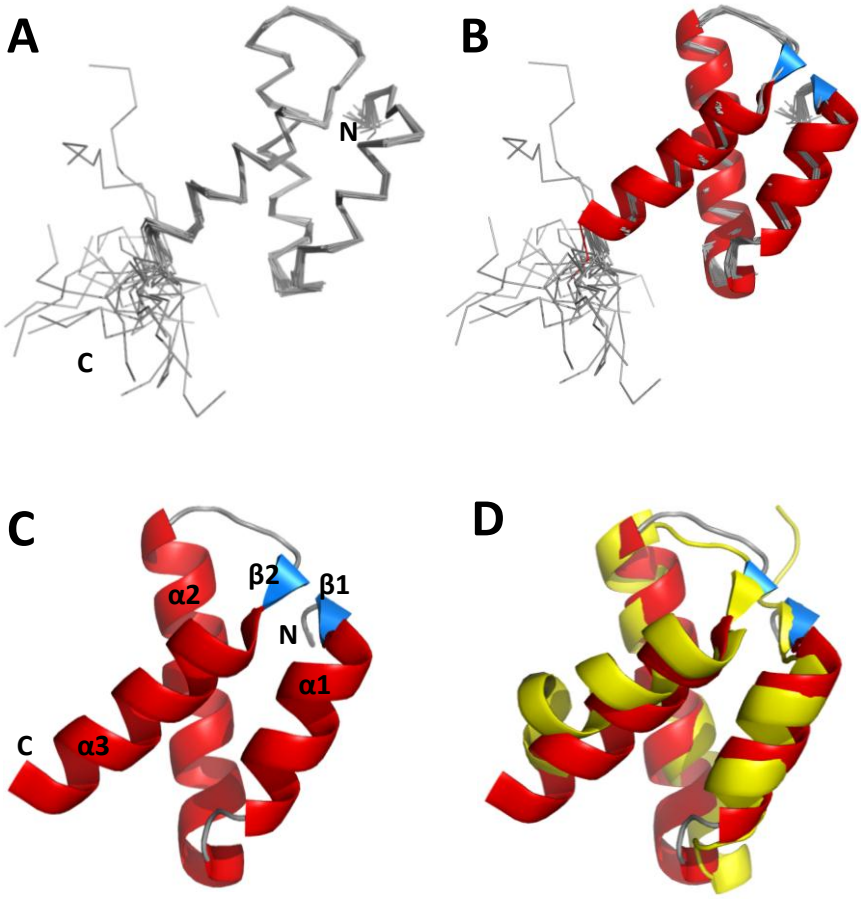

Figure 4.14 Solution NMR structure of Asr4154 from *Anabaena* sp. PCC 7120 (PDB ID 2L09) A) Ensemble of the 20 lowest energy models deposited to the Protein Data Bank, shown in a ribbon configuration, hiding the side chain and hydrogen atoms for clarity, and with labeled N- and C-termini. Residues 1-62 are shown. B) Ribbon ensemble with the lowest energy model overlayed and shown in a cartoon configuration with colored secondary structure. C) Cartoon representation of the lowest energy model of Asr4154 with labeled secondary structure elements in the order β1-α1- α2- β2- α3. D) Asr4154 lowest energy structure aligned with the lowest energy model for the solution NMR structure of the C-terminal fragment of BchB from *Chlorobium tepidum* (ctBchB-C) in yellow (PDB ID 2KRU). Unless otherwise noted, residues 1-54 are shown for Asr4154 and residues 484-537 for ctBchB-C.

The surface charge of Asr4154 can be divided into a highly electropositive face and a highly electronegative face, separated by a small ring of neutral charge, as visualized with the APBS tool (5) in PyMOL (58) (Figure 4.15). Lysine residues K10, K14, and K25 together with arginine residues R4, R21, and R49 contribute to the electropositive composition on the surface containing strand β1 and helices α1 and α2. The opposite face contains the acidic residues aspartate D38 and glutamates E36, E43, E46, and E51, which contribute to the electronegative potential on the surface containing strand β2 and helices α2 and α3. The sequence conservation between Asr4154 and ctBchB-C is made up of both polar and nonpolar residues (Figure 4.16), including three aromatic amino acids: tryptophan W5, and phenylalanines F18 and F19 that reside near the electropositive surface of Asr4154. A ConSurf analysis was used to further assess the degree of sequence conservation for surface residues across the Pfam. The top 50 protein BLAST sequences to Asr4154 from the KEGG database (38) were used to generate a multiple sequence alignment file, which was supplied to the ConSurf server (3) to generate a color-coded model for Asr4154 (Figure 4.17). A highly conserved patch was located at the surface consisting of the C-terminal helix, α3, and the region between helices α1 and α2. Here, many polar amino acids were found to be highly conserved including residues K12, R21, R27, R49, N15, and Q23. Additionally, proline P17 and phenylalanines F19 and F52 were located in this highly conserved surface patch. The opposite side of the protein, containing the N-terminus, anti-parallel strands β2 and β3, and the N-terminal ends of helices α1, α2, and α3, was highly variable in sequence across the represented sequences from the protein family.

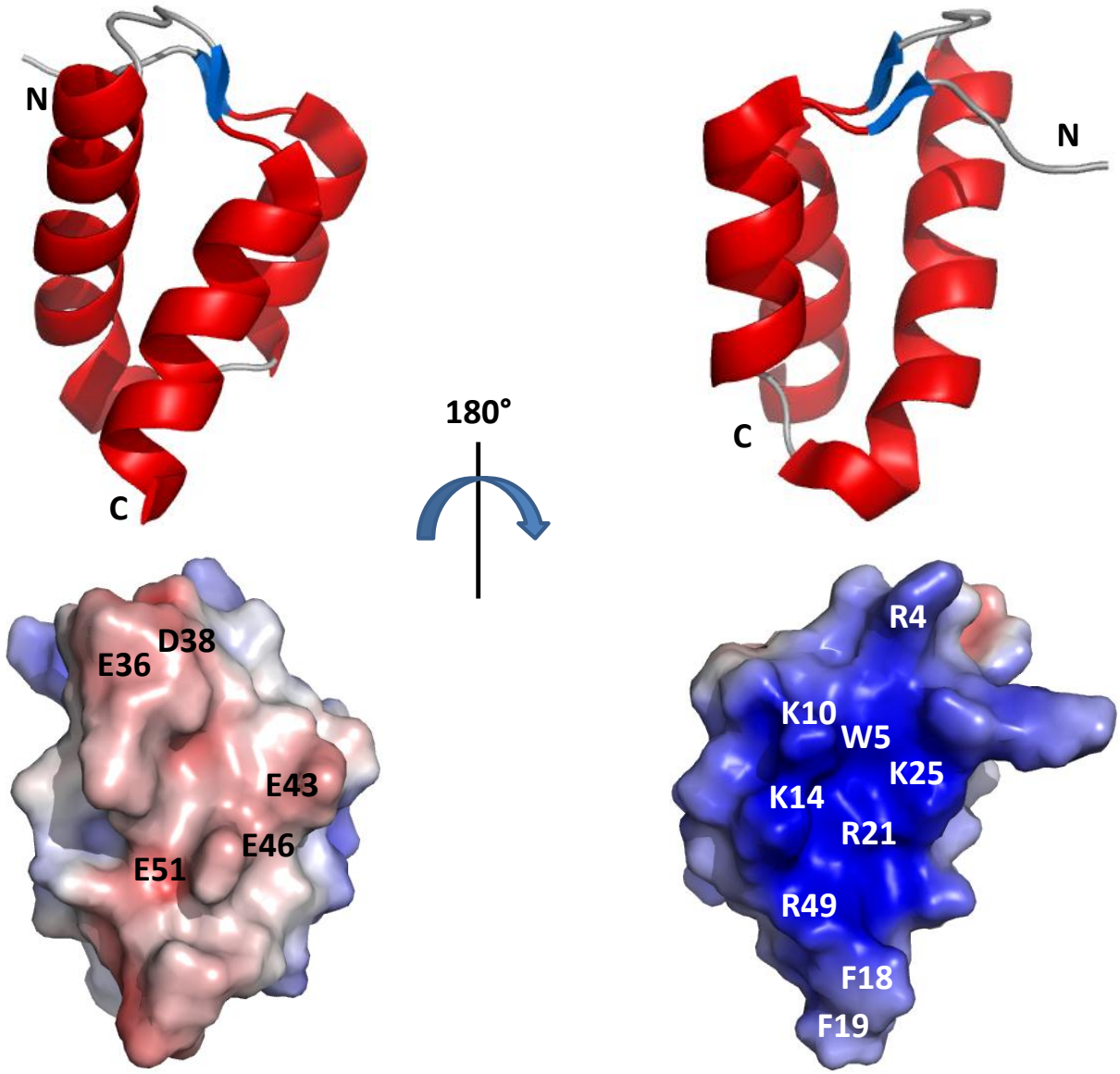

Figure 4.15 Electrostatic surface potential map generated using APBS (5). The negatively charged surface regions are shown in red, positively charged shown in blue, and neutral shown in white (±5 kT/e). Shown in the right-hand panel is a 180° rotation of the left-hand figures to show the opposite face of the protein surface. Asr4154 residues 1-54 are shown with specific acidic, basic, and aromatic amino acids indicated.

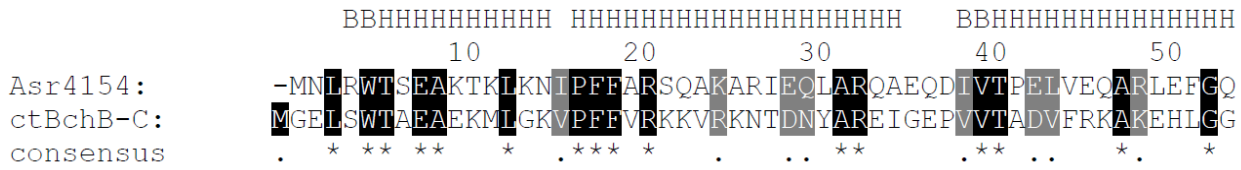

Figure 4.16 Multiple sequence alignment of Asr4154 from *Anabaena* sp. PCC 7120 and the C-terminal fragment of BchB from *Chlorobium tepidum* (ctBchB-C) generated from ClustalOmega and BOXSHADE. The corresponding secondary structure elements from Asr4154 are labeled above the residue numbers in the sequence. The ctBchB-C fragment of residues 484-537 is shown with the addition of a non-native N-terminal methionine residue.

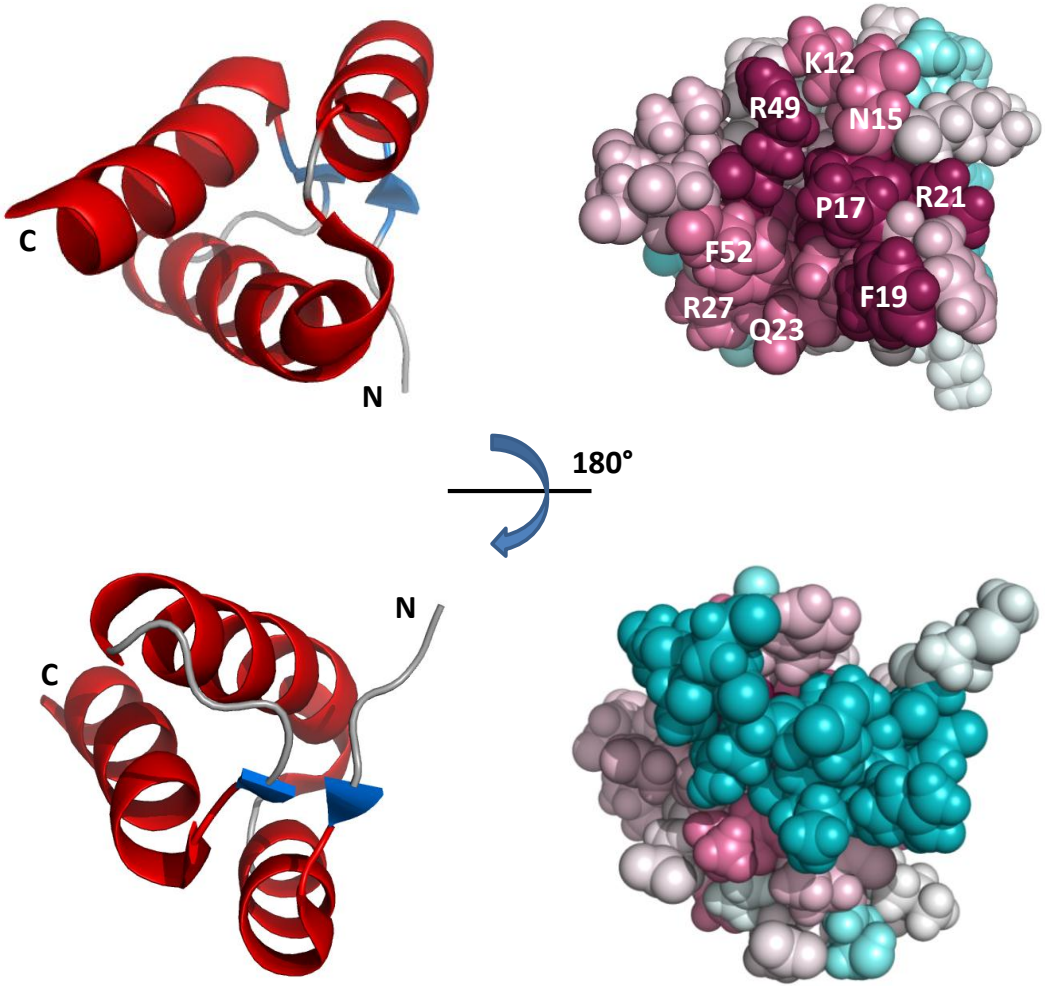

Figure 4.17 ConSurf (3) analysis of Asr4154 using the top 50 BLAST hits using the KEGG database (38) (*E*-values range from 2 x 10^{-27} to 5 x 10^{-2}). The top panel indicates a highly conserved surface patch (magenta and pink) predominantly composed of electropositive residues (K12, R21, R27, and R49) but also includes aromatic residues F19 and F52, asparagine N15, glutamine Q23, and proline P17. Shown in the lower panel is a 180° rotation of the top figures to show the opposite surface has high sequence variability indicated by a mostly cyan-colored patch. Asr4154 residues 1-54 are shown and the right hand figures are shown in the spheres configuration.

The Pfam PF08369 is found in 288 sequenced organisms, mostly represented in bacteria where it is putatively involved in the biosynthesis of bacteriochlorophylls (Bch) but also found in plant chloroplasts involved in the synthesis of plant chlorophylls (Chl). This family is annotated as the 57 kD subunit of the light-independent protochlorophyllide (Pchlide) reductase enzymes, also known as dark-operative Pchlide oxidoreductases (DPOR). The DPOR protein is a nitrogenase-like enzyme composed of an L-protein domain (BchL homodimer) and an NB-protein domain (BchN-BchB heterotetramer) which together, carries out the rate-limiting step of the Bch/Chl biosynthesis pathway where the C17=C18 double bond of Pchlide is reduced to form chlorophyllide *a*, the direct precursor of the light-harvesting molecule chlorophyll *a* (9, 24, 25, 47, 57). Crystal structures have recently been released of the BchL L-protein from *Rhodobacter sphaeroides* (PDB ID 3FWY) (57) and the BchN-BchB heterotetramer from *Rhodobacter capsulatus* (PDB ID 3AEK) (47) and *Thermosynechococcus elongatus* (PDB ID 2XDQ) (8), and have provided important structural details of both the similarities and differences between their respective nitrogenase homologs (nitrogenase Fe protein is homologous to the BchL protein; nitrogenase MoFe protein homologous to the BchN-BchB heterotetramer). Asr4154 from *Anabaena*, like several representatives from Pfam PF08369, is annotated as an independent protein. However, there are also many PF08369 members that are found on the C-terminal end of BchB proteins (Pfam PF00148), like ctBchB-C from *Chlorobium tepidum*. The two BchB proteins for which crystal structures are available, share ~30% identity with Asr4154 for the C-terminal domains that align (Figure 4.18). Both crystal structures however, were solved as truncations lacking the C-terminal domains that align with Asr4154, and it should also be noted that the homologous nitrogenase MoFe proteins typically lack this same C-terminal component. For BchB/ChlB proteins though, the C-terminal domain is highly conserved. The C-terminal domain of the BchB protein from *Rhodobacter capsulatus* (residues 475-520), is predicted to contain a disordered region important for the distinctive character of DPOR function (47). Therefore, Asr4154 and ctBchB-C represent intriguing novel structures for this family of proteins.

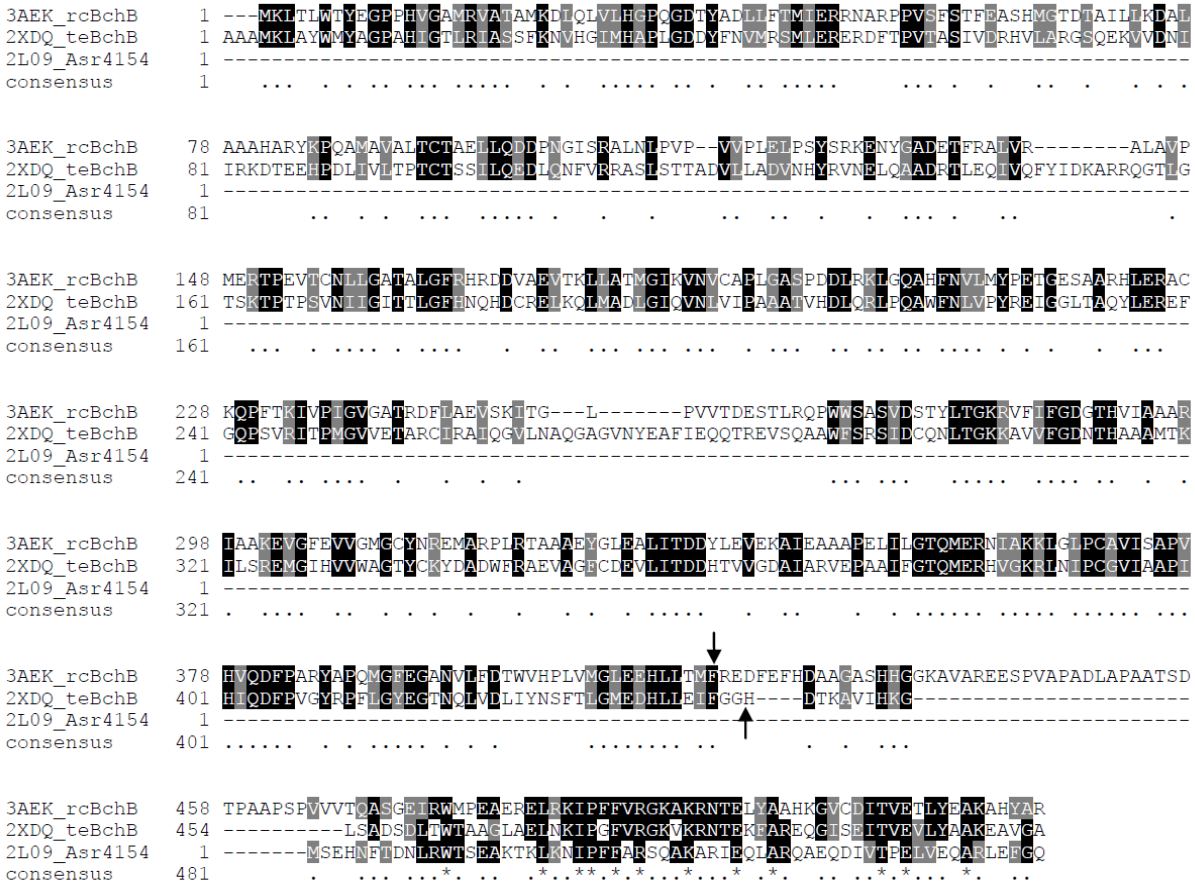

Figure 4.18 Multiple sequence alignment of Asr4154 from *Anabaena* sp. PCC 7120 and BchB proteins from *Rhodobacter capsulatus* (rcBchB) and *Thermosynechococcus elongatus* for which the crystal structures are available (PDB IDs 3AEK and 2XDQ respectively). The specific location of the C-terminal truncation for each BchB crystal structure is indicated with an arrow for the corresponding sequence. The alignment was generated using ClustalOmega and BOXSHADE.

4.5 Introduction to NESG target ID: LaR80A

In this section, the solution NMR structure of a mucin-binding domain (MucBP) domain (fragment of residues 182-294) of the protein LBA1460 from *Lactobacillus acidophilus* is described. This MucBP protein fragment is a member of Pfam PF06458, a family of mucin-binding adhesion proteins (adhesins), mostly represented in bacteria (including human pathogens), and located on the cell surface. The MucBP proteins are responsible for facilitating cellular attachment to a host organism by adhering to mucus secreted from the host's epithelial cells (17). Mucus is largely composed of mucin (MUC) proteins, a family of high molecular weight oligomers, of which there are 17 members in humans. In the human intestine, mucins are found as either membrane bound (ex: MUC1, MUC3, MUC4, MUC12, MUC13, and MUC17) or secreted (MUC2, MUC5B, MUC5AC, and MUC6) proteins and are heavily *O*-glycosylated. The architecture of *O*-glycosylations on mucins is quite diverse, found as both linear and branching *O*-glycans, of which the identities and arrangement of the sugar moieties are thought to contribute to the specific activity of the particular mucin (36). The mucin-binding MucBP domains share sequence homology with the mucus-binding proteins (MUBs) and the *Listeria monocytogenes* internalin (Inl) proteins. MucBP domains are often found as repeats within a much larger protein sequence, as is the case within many MUB and Inl proteins. The LBA1460 protein from *L. acidophilus* however, contains a single MucBP domain, similar to the Spr1345 protein from the infamous human pathogen *Streptococcus pneumonia*. The MucBP from Spr1345 was recently solved by X-ray crystallography (PDB ID 3NZ3) and was shown to have a ubiquitin-like β-sandwich fold (17) similar to that of the crystal structure of the Mub2 domain of the fifth MucBP repeat of MUB (MubR5) from *Lactobacillus reuteri* (PDB ID 3I57) (45), and exhibited structural similarities to the internalin B (InlB) B-repeat domain from *Lactobacillus monocytogenes* (PDB ID 2Y5P) (18) and an adhesion exoprotein from *Pediococcus pentosaceus* (PDB ID 2KYW). The solution NMR structure of MucBP from LBA1460 described here, also adopts an immunoglobulin-like β-sandwich fold and exhibits many close structural similarities to Spr1345 and MubR5. The presented research will

contribute to the understanding of Pfam PF06458 and other bacterial adherence proteins, with hopes in achieving enhanced knowledge of how pathogens use surface-exposed adhesins, in order to potentially design new antibiotics or protein-based vaccines.

4.5.1 Materials and Methods

4.5.1.1 Expression and purification of MucBP samples for NMR

The gene fragment coding for the putative mucin-binding domain (MucBP) of the protein LBA1460 from *L. acidophilus* (UniProtKB/Swiss-Prot ID, Q5FJ43_LACAC) was cloned, expressed, and purified according to standard protocols of the NESG consortium to prepare protein samples for NMR (1). Briefly, the gene fragment (encoding for residues 182-294) was cloned into a pET21 expression vector with a non-native N-terminal methionine and C-terminal His_6 affinity tag (LEHHHHHH), transformed in *E. coli* BL21 (DE3) pMGK cells, and grown at 37 °C in MJ9 minimal media until reaching an OD_{600} of 0.6, then equilibrated to 17 °C and induced with 1 mM isopropyl-β-*D*-thiogalactopyranoside (IPTG) before harvesting the cell pellet by centrifugation. To generate isotopically-labeled [*U*-^{13}C,^{15}N]-MucBP (NC) and uniform 100% ^{15}N, 5% biosynthetically directed ^{13}C (NC5) MucBP for NMR, cells were grown with $(U\text{-}^{15}NH_4)_2SO_4$ and *U*-^{13}C glucose as the sole sources of nitrogen and carbon. Cells were harvested by centrifugation, followed by lysis using sonication, and then subject to high speed centrifugation to recover the soluble fraction. MucBP protein was first purified with a Ni-affinity column (HisTrap HP, 5 mL) using the buffer conditions 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 40 mM imidazole, 1 mM TCEP, followed by elution in identical buffer with 500 mM imidazole. The MucBP solution was then exchanged into buffer for gel-filtration chromatography containing 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM MES, pH 6.5, and purified on a HiLoad 26/60 Superdex 75 column using an ÄKTAxpress™ system (GE Healthcare). Sample purity of labeled NC- and NC5-MucBP protein was assessed using SDS PAGE (>95%

purity) and MALDI-TOF mass spectrometry to verify the molecular weight (14.6 kDa for NC and 14.2 kDa for NC5 samples). Samples were concentrated to 1.2 mM by centrifugation and then supplemented with 10% v/v D_2O to the final solution for NMR studies.

4.5.1.2 NMR spectroscopy and structure calculations

NC- and NC5-MucBP samples were used to collect NMR data at 20 °C in 5 mm Shigemi tubes on a 850 MHz Bruker Avance III spectrometer and a 600 MHz Varian Inova spectrometer. MucBP behaved as a monomeric protein according to an isotropic overall rotational correlation time (τ_c) of 8.7 ns, estimated from ^{15}N T_1 and $T_{1\rho}$ relaxation data (not shown) under the conditions used for NMR spectroscopy (1.2 mM MucBP protein, 10% v/v D_2O, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM MES, pH 6.5). The D_2O-exchanged samples were prepared by freezing NC-MucBP in liquid nitrogen, followed by overnight lyophilization and resuspension into 99.9% D_2O (Acros Organics). NMR spectra were processed using NMRPipe (15) and visualized with Sparky 3.110 (27). Chemical shift resonance assignments for the peptide backbone and side chain nuclei were made using the experiments: ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) (Figure 4.19) and ^{1}H-^{13}C HSQC, HNCA, HNCO, HN(CO)CA, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(C)CH-TOCSY, H(C)CH-COSY, H(CC)(CO)NH-TOCSY, (H)CC(CO)NH-TOCSY, ^{15}N-edited NOESY-HSQC (τ_m = 70 ms), ^{13}C-edited NOESY-HSQC (τ_m = 70 ms) experiment optimized for aromatic carbons and a non-uniform sampling (NUS) experiment optimized for aliphatic carbons, followed by additional 3D (H)CCH-TOCSY and 4D ^{13}C-^{13}C-HMQC-NOESY-HMQC (τ_m = 70 ms) experiments collected on D_2O-exchanged NC-MucBP samples. A high resolution ^{13}C-HSQC constant time experiment collected on a NC5-MucBP sample was collected for assigning stereospecific methyl group resonances of Val and Leu isopropyl side chains. Pulse sequences for Varian 2D and 3D experiments came from the Varian BioPack library and the 4D NOESY came from Lewis Kay (University of Toronto). The

TopSpin standard library was used for Bruker pulse sequences. Assignments for the amide backbone $^{1}H^{N}$, $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$ and side chain $^{13}C^{\beta}$ and $^{1}H^{\beta}$ resonances were made after manual peak picking and subsequent submission to the PINE server (NMRFAM) (4) for auto-assignment. Additional side chain resonances were completed manually. Final chemical shifts, NOESY peak lists, and raw FIDS were deposited in the BioMagResDB (BMRB accession number, 17754).

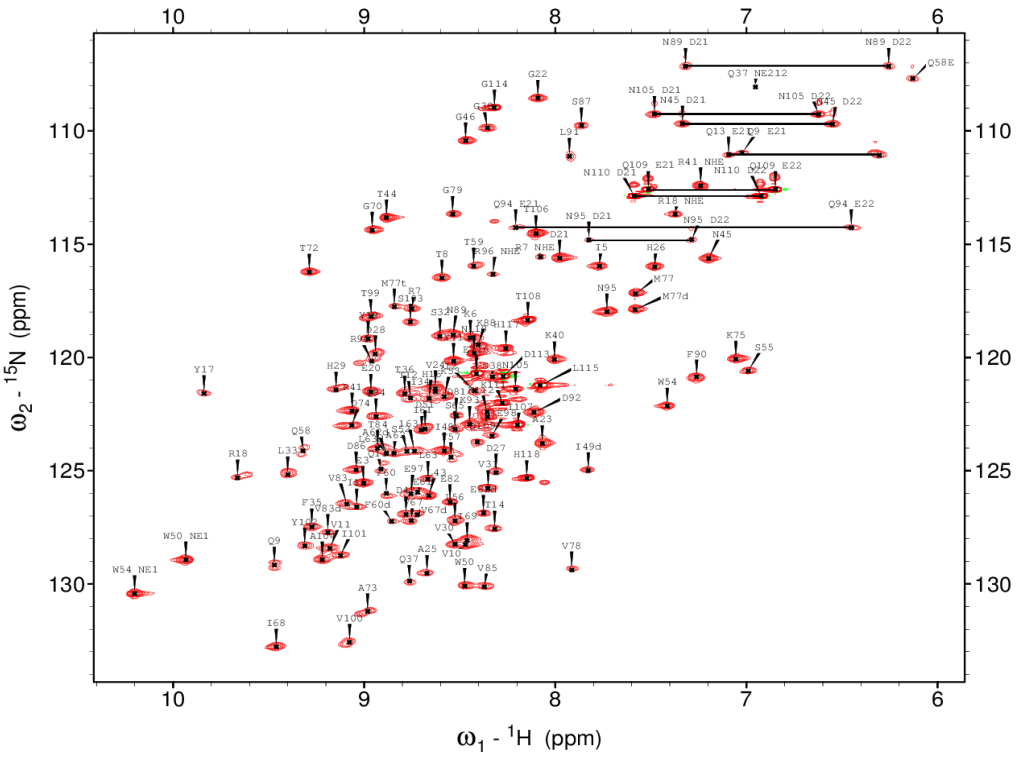

Figure 4.19 Two dimensional ^{1}H-^{15}N HSQC spectrum of 1.2 mM for [*U*-^{13}C, ^{15}N]-MucBP in a 10% v/v D_2O solution containing: 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3, 20 mM MES, pH 6.5, acquired on a Bruker AVANCE III 850 MHz spectrometer at 20 °C. Backbone and side chain resonance assignments are indicated above with their corresponding one-letter amino acid codes followed by their sequence number (original 1-122 naming scheme for assignments, corresponding to Met1 followed by residues 182-294 in the LBA1460 sequence plus an additional LEHHHHHH C-terminal affinity tag) and H-N pair classification.

The solution NMR structures of MucBP were calculated with CYANA 2.1 (28, 30) to automatically assign NOEs while using dihedral angle constraints for φ and ψ dihedral angles derived using Talos+ (61), manually derived hydrogen bonds constraints, and two sets of amide backbone one-bond ^{1}H-^{15}N residual dipolar coupling (RDC) constraints derived by comparing RDC data collected for isotropic and two partially aligned samples in NMR buffer of 1.1 mM NC5-MucBP using NH *J*-modulation experiments to measure one-bond ^{1}H-^{15}N couplings (65). The first of these samples was aligned in a solution of 12.5 mg/mL Pf1 filamentous bacteriophage and the second sample was aligned in a neutral stretched 5% polyacrylamide gel prepared in 19:1 mixture of acrylamide and *N*,*N*′-methylenebisacrylamide as described previously (29, 44, 56). All constraint files were then converted to XPLOR / CNS format using PdbStat (6) with an increase of 10% to the upper bounds for NOEs, and then the final structures were calculated with XPLOR-NIH-2.20 (59), following refinement in explicit water using CNS version 1.1 (10). Structure ensembles and constraints were submitted to the PSVS server version 1.4 (6) to examine NOE assignments and further identify flagrant NOE violations which were evaluated manually. NOESY peak lists and resonance assignments were refined using NMR RPF (33) quality assessment scores to check the goodness-of-fit between the XPLOR-calculated structures and the NOESY peak lists. Final constraints and ensemble of 20 structures sorted by lowest energy were submitted to the Protein Data Bank (PDB ID 2LFI).

4.5.2 Results and Discussion

The solution NMR structure of LBA1460_MucBP is presented in Figure 4.20 and the corresponding structural statistics are found in Table 4.5. The structure adopts an immunoglobulin-like β-sandwich fold that most closely resembles the cylindrical fold of the immunoglobulin-binding superfamily of proteins (48). The MucBP structure is composed of seven β-strands, β1 (residues 182-199), β2 (204-222), β3 (228-230), β4 (235-240), β5 (251-253), β6 (261-266), β7 (277-284), predominantly formed by anti-parallel β-sheet, however a short region of parallel β-sheet is adopted between residues

277-281 of strand β7 and residues 194-199 of strand β1. A long loop segment is also present between strands β4-5 and β5-6, the location of four proline residues. The solution NMR ensemble is in strong agreement with the X-ray crystal structure, solved simultaneously by the NESG (PDB ID 3Q69). For residues in ordered secondary structure (residues 182-199, 204-222, 228-230, 235-240, 251-253, 261-266, and 277-284), the average RMSD for backbone atoms is 1.6 Å, and the structural alignment is in Figure 4.20B. One subtle difference between the NMR ensemble and the crystal structure is the outward shift in position of the C-terminal end of strand β2 in the NMR structure. Even though the NMR structure was calculated using the same hydrogen bond restraints as the crystal structure in this region, which were verified by cross strand $^{1}H_{\alpha}$-$^{1}H_{\alpha}$, and $^{1}H_{\alpha}$-$^{1}H_{N}$, and $^{1}H_{N}$-$^{1}H_{N}$ NOEs, the side chain resonances of two surface histidine residues, His206 and His209, were not able to be assigned. These surface residues likely have considerable flexibility, making the chemical shifts of the side chain atoms nearly invisible, and contributing to the subtle bow of the peptide backbone in the model.

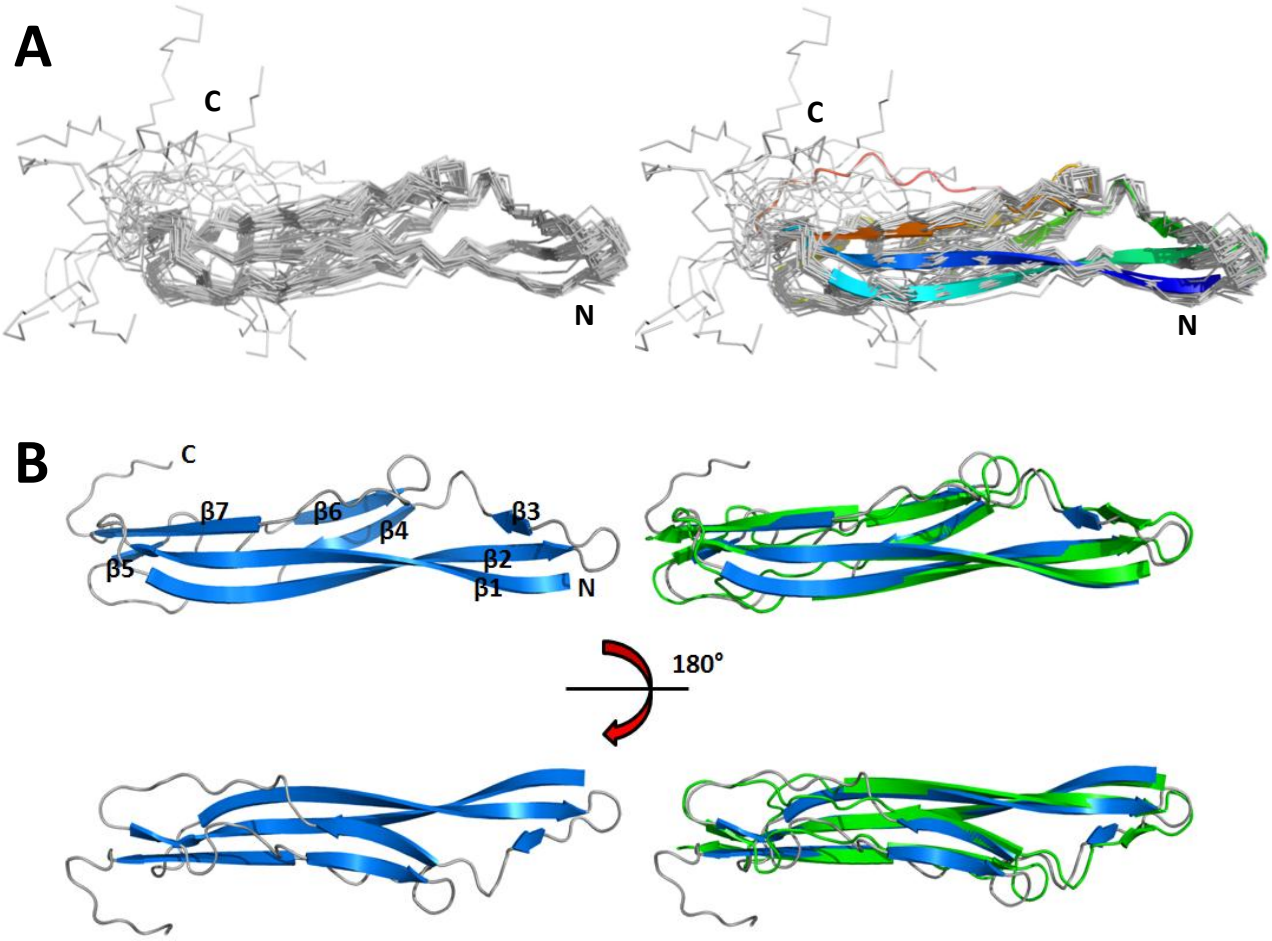

Figure 4.20 Solution NMR structure for the MucBP domain (fragment 182-294) of LBA1460 from *L. acidophilus* (PDB ID 2LFI). A) Shown in a ribbon representation, is the ensemble of 20 lowest energy models for residues 182-302 which includes the disordered His_6 (LEHHHHHH) affinity tag at the C-terminus. The right-hand panel shows a cartoon representation of the lowest energy model aligned with the ensemble and colored blue to red in N-C directionality. B) Lowest energy model of the MucBP solution NMR ensemble (gray and blue) aligned with the X-ray crystal structure (green) of MucBP solved by the NESG (PDB ID 3Q69). Residues 182-294 are shown for 2LFI and residues 182-285 for 3Q69.

Completeness of resonance assignments[a]	
Backbone (%)	99.1
Side chain (%)	93.4
Aromatic (%)	87.2
Stereospecific methyl (%)	95.0
Conformationally-restricting constraints[b]	
Distance constraints	
Total	2031
Intra-residue [$i = j$]	360
Sequential [$\| i - j \| = 1$]	663
Medium range [$1 < \| i - j \| < 5$]	246
Long range [$\| i - j \| \geq 5$]	762
Dihedral angle constraints	157
Hydrogen bond constraints	78
NH RDC constraints (Phage/Gel)	31/31
Number of constraints per residue	19.2
Number of long range constraints per residue	7.0
Residual constraint violations[b]	
Average number of distance violations per structure	
0.1 - 0.2 Å	23.9
0.2 – 0.5 Å	7.0
> 0.5 Å	0.0
Average number of distance violations per structure	
1 - 10°	12.5
> 10°	0.0
RMSD from average coordinates (Å) [b,d]	
Backbone atoms	1.1

Heavy atoms			1.4
Procheck Ramachandran statistics [b,d]			
Most favored regions (%)			96.8
Allowed regions (%)			3.1
Disallowed regions (%)			0.1
Global quality scores (Raw/ *Z*-score) [b]			
Verify3D	0.2	/	-3.5
ProsaII	0.1	/	-2.2
Procheck G-factor (phi-psi)[d]	-0.5	/	-1.8
Procheck G-factor (all)[d]	-0.4	/	-2.5
MolProbity Clashscore	15.1	/	-1.1
RPF Scores[c]			
Recall / Precision	0.91		0.88
F-measure / DP-score	0.89		0.76

[a] Refers to chemical shifts for residues 182-294.
[b] Calculated for the ensemble of 20 structures using PSVS version 1.4 (6). Average distance violations were calculated using the sum over r^{-6}.
[c] RPF scores (33) calculated for the ensemble of 20 structures reflecting the goodness-of-fit to the NOESY data and resonance assignments.
[d] Ordered residue ranges: 182-206, 209-241, 245-258, 262-272, 277-288 with the sum of φ and ψ order parameters > 1.8.

Table 4.5 Summary of solution NMR structure statistics for the MucBP domain of LBA1460 from *L. acidophilus* (PDB ID 2LFI).

Using the atomic coordinates for the solution NMR ensemble as input, a DALI (31) search was used to identify proteins with similar structures to MucBP. After the X-ray crystal structure of the MucBP from LBA1460 (*Z*-score 11.2), the X-ray crystal structure of the MucBP domain of Spr1345 from *Streptococcus pneumoniae* was identified as the next closest structure (*Z*-score 8.3), followed by the MubR5 X-ray crystal structure (*Z*-score 7.8). All structures adopt the characteristic immunoglobulin-like β-sandwich, first identified by MacKenzie *et al*. for the MUB protein from *L. reuteri* (45). A common feature among these protein domains is the compact hydrophobic core, responsible for the narrow cylindrical fold. The core of the MucBP domain contains several hydrophobic residues that are highly conserved based on a ConSurf (3) analysis (Figures 4.21 and 4.22) using the top 50 homologous sequences to the LBA1460 MucBP domain, identified from a BLAST search provided by the KEGG database (38).. This was also noted by Du *et al*. in the crystal structure report of the Spr1345 MucBP from *S. pneumoniae*, where an alignment of several MucBP domains from different species showed a high degree of sequence conservation for many core hydrophobic residues and are largely responsible for the compact nature of the cylindrical fold (17). MacKenzie *et al.* identified in the initial report of the MubR5 crystal structure that core aromatic residues Trp135 and Trp138 stabilize a cap on the end of the cylindrical structure of the B2 domain together with an unusually-positioned arginine side chain (Arg122) between the two indole rings in the core (45). These tryptophan residues are structurally homologous with Trp230 and Trp234 in the MucBP domain of LBA1460; whereas a core glutamine (Gln217) is positioned similarly in LBA1460 to that of Arg122 in MubR5 (Figure 4.22A). In addition, the N-terminal cap from both proteins also contains a structurally conserved core phenylalanine residue (Phe215 in LBA1460; Phe120 in MubR5) which, in the LBA1460 MucBP solution NMR structure, has several long-range NOEs to glutamines Gln238 and Gln274, valines Val263 and Val265, and leucine Leu236. The C-terminal cap of the structure also contains structurally conserved core aromatics, tyrosines Tyr197, Tyr199, and Tyr282 in the LBA1460 (corresponding to Tyr102, Tyr104, and Tyr181 in MubR5), which appear to stabilize the opposite end of

the cylinder (Figure 4.22B). The loop segment, between strands β4-5 and β5-6 is also a feature shared between Spr1345 and MubR5. In LBA1460, this region contains four proline residues and several hydrophobic amino acids that have many long range NOEs to residues across the barrel of the cylinder to strands β1, β2, and β7. Altogether, the core of the MucBP domain appears to require these hydrophobic and aromatic residues in order to tie down the remainder of the fold into a compressed cylinder.

In summary, the solution NMR structure of the MucBP domain for the LBA1460 protein from *L. acidophilus* has been presented. To date, the LBA1460 protein has not been functionally characterized. It is likely that the MucBP domain of LBA1460 has mucin-binding activity similar to that recently observed in Spr1345 (12, 17) from the human pathogen, *S. pneumoniae*. Bumbaca *et al*. observed that the MucBP of Spr135 recognized specifically the carbohydrate moiety of the mucins tested, and that the mucin identity was less important than the sugar-identity for binding (12). However, it should be noted that many mucins are coated with a variety of different oligosaccharides, and with different linkages (although often *O*-linked) (13), making the identification of a specific mucin-binding site on MucBP proteins not necessarily obvious. Du *et al*. though made the observation that a few surface residues in the C-terminal cap of Spr1345 are highly conserved across MucBP proteins, are could serve in carbohydrate recognition (17). In LBA1460, a surface tyrosine, Tyr251, is homologous to Tyr73 of Spr1345 in this region, and could potentially serve such a role. Any specific residues required for mucin-binding, will require additional *in vitro* and *in vivo* experiments, to complement a better understanding of how bacteria use adhesin proteins to adhere to and colonize with their host cells.

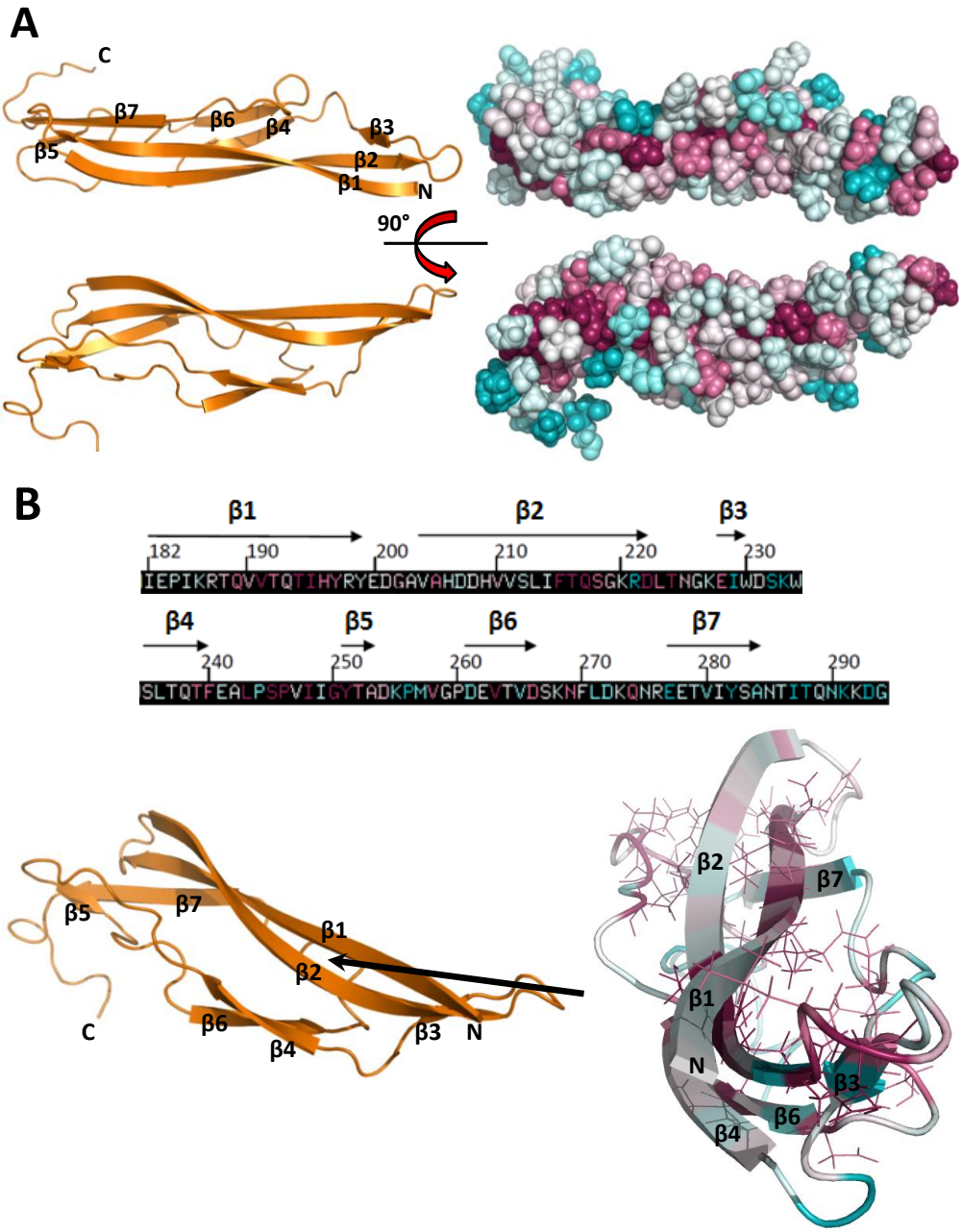

Figure 4.21 ConSurf (3) structure of MucBP from LBA1460 shown in A) sphere representation and also in B) cartoon representation with highly conserved residue side chains in line representation highlighting the conservation of core residues. The viewing direction down the N-terminal side of the cylinder in the right-hand structure is depicted with an arrow in the left-hand structure.

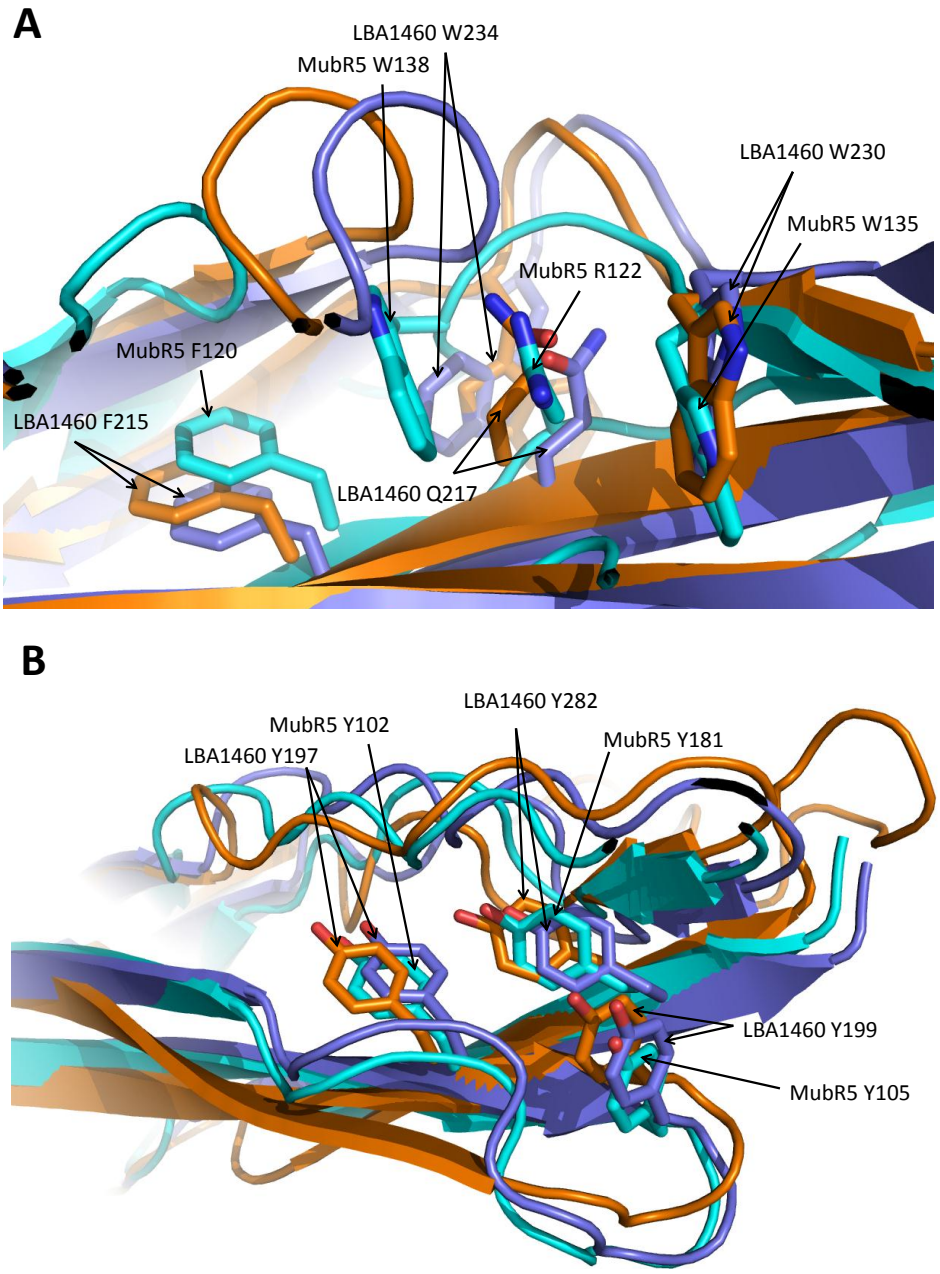

Figure 4.22 Structural alignment of the MucBP domain of LBA1460 from *L. acidophilus* (PDB ID 2LFI for lowest energy NMR model in orange; PDB ID 3Q69 for X-ray crystal structure in purple) with the B2 domain of MubR5 from *L. reuteri* (PDB ID 3I57 in cyan) showing the position of core aromatic residues at the A) N-terminal cap of the cylinder (including residue Arg122 for MubR5 and Gln217 for LBA1460) and B) C-terminal cap of the cylinder.

4.6 References

1. Acton, T., R. Xiao, S. Anderson, J. Aramini, W. Buchwald, C. Ciccosanti, K. Conover, J. Everett, K. Hamilton, Y. Huang, H. Janjua, G. Kornhaber, J. Lau, D. Lee, G. Liu, M. Maglaqui, L. Ma, L. Mao, D. Patel, P. Rossi, S. Sahdev, R. Shastry, S. GVT., Y. Tang, S. Tong, D. Wang, H. Wang, L. Zhao, and G. Montelione. 2010. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods in Enzymology 493:21-60.
2. Arteni, A. A., P. Zhang, N. Battchikova, T. Ogawa, E. M. Aro, and E. J. Boekema. 2006. Structural characterization of NDH-1 complexes of Thermosynechococcus elongatus by single particle electron microscopy. Biochim Biophys Acta 1757:1469-1475.
3. Ashkenazy, H., E. Erez, E. Martz, T. Pupko, and N. Ben-Tal. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529-533.
4. Bahrami, A., A. H. Assadi, J. L. Markley, and H. R. Eghbalnia. 2009. Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput Biol 5:e1000307.
5. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98:10037-10041.
6. Battacharya, A., R. Tejero, and G. Montelione. 2007. Evaluating protein structures determined by structural genomics consotria. Proteins 66:778-795.
7. Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of Escherichia coli K-12. Science 277:1453-1462.
8. Bröcker, M. J., S. Schomburg, D. W. Heinz, D. Jahn, W. D. Schubert, and J. Moser. 2010. Crystal structure of the nitrogenase-like dark operative

protochlorophyllide oxidoreductase catalytic complex (ChlN/ChlB)2. J Biol Chem 285:27336-27345.

9. Bröcker, M. J., D. Wätzlich, M. Saggu, F. Lendzian, J. Moser, and D. Jahn. 2010. Biosynthesis of (bacterio)chlorophylls: ATP-dependent transient subunit interaction and electron transfer of dark operative protochlorophyllide oxidoreductase. J Biol Chem 285:8268-8277.
10. Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905-921.
11. Buell, C. R., V. Joardar, M. Lindeberg, J. Selengut, I. T. Paulsen, M. L. Gwinn, R. J. Dodson, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, S. Daugherty, L. Brinkac, M. J. Beanan, D. H. Haft, W. C. Nelson, T. Davidsen, N. Zafar, L. Zhou, J. Liu, Q. Yuan, H. Khouri, N. Fedorova, B. Tran, D. Russell, K. Berry, T. Utterback, S. E. Van Aken, T. V. Feldblyum, M. D'Ascenzo, W. L. Deng, A. R. Ramos, J. R. Alfano, S. Cartinhour, A. K. Chatterjee, T. P. Delaney, S. G. Lazarowitz, G. B. Martin, D. J. Schneider, X. Tang, C. L. Bender, O. White, C. M. Fraser, and A. Collmer. 2003. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. Proc Natl Acad Sci U S A 100:10181-10186.
12. Bumbaca, D., J. E. Littlejohn, H. Nayakanti, A. H. Lucas, D. J. Rigden, M. Y. Galperin, and M. J. Jedrzejas. 2007. Genome-based identification and characterization of a putative mucin-binding protein from the surface of Streptococcus pneumoniae. Proteins 66:547-558.
13. Byrd, J. C., and R. S. Bresalier. 2004. Mucins and mucin binding proteins in colorectal cancer. Cancer Metastasis Rev 23:77-99.

14. Cornilescu, G., F. Delaglio, and A. Bax. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289-302.
15. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277-293.
16. Dessailly, B. H., R. Nair, L. Jaroszewski, J. E. Fajardo, A. Kouranov, D. Lee, A. Fiser, A. Godzik, B. Rost, and C. Orengo. 2009. PSI-2: structural genomics to cover protein domain family space. Structure 17:869-881.
17. Du, Y., Y. X. He, Z. Y. Zhang, Y. H. Yang, W. W. Shi, C. Frolet, A. M. Di Guilmi, T. Vernet, C. Z. Zhou, and Y. Chen. 2011. Crystal structure of the mucin-binding domain of Spr1345 from Streptococcus pneumoniae. J Struct Biol 174:252-257.
18. Ebbes, M., W. M. Bleymüller, M. Cernescu, R. Nölker, B. Brutschy, and H. H. Niemann. 2011. Fold and function of the InlB B-repeat. J Biol Chem 286:15496-15506.
19. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126-2132.
20. Finnin, M. S., M. P. Cicero, C. Davies, S. J. Porter, S. W. White, and K. N. Kreuzer. 1997. The activation domain of the MotA transcription factor from bacteriophage T4. EMBO J 16:1992-2003.
21. Finnin, M. S., D. W. Hoffman, and S. W. White. 1994. The DNA-binding domain of the MotA transcription factor from bacteriophage T4 shows structural similarity to the TATA-binding protein. Proc Natl Acad Sci U S A 91:10972-10976.
22. Frickey, T., and A. Lupas. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20:3702-3704.

23. Friedrich, T., and H. Weiss. 1997. Modular evolution of the respiratory NADH:ubiquinone oxidoreductase and the origin of its modules. J Theor Biol 187:529-540.
24. Fujita, Y. 1996. Protochlorophyllide reduction: a key step in the greening of plants. Plant Cell Physiol 37:411-421.
25. Fujita, Y., and C. E. Bauer. 2000. Reconstitution of light-independent protochlorophyllide reductase from purified bchl and BchN-BchB subunits. In vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. J Biol Chem 275:23583-23588.
26. Gabanyi, M. J., P. D. Adams, K. Arnold, L. Bordoli, L. G. Carter, J. Flippen-Andersen, L. Gifford, J. Haas, A. Kouranov, W. A. McLaughlin, D. I. Micallef, W. Minor, R. Shah, T. Schwede, Y. P. Tao, J. D. Westbrook, M. Zimmerman, and H. M. Berman. 2011. The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods. J Struct Funct Genomics 12:45-54.
27. Goddard, T., and D. Kneller. SPARKY 3, University of California, San Francisco.
28. Güntert, P. 2004. Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353-378.
29. Hansen, M. R., L. Mueller, and A. Pardi. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5:1065-1074.
30. Herrmann, T., P. Güntert, and K. Wüthrich. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209-227.
31. Holm, L., and P. Rosenström. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545-549.
32. Huang, Y. J., H. N. Moseley, M. C. Baran, C. Arrowsmith, R. Powers, R. Tejero, T. Szyperski, and G. T. Montelione. 2005. An integrated platform for automated analysis of protein NMR structures. Methods Enzymol 394:111-141.

33. Huang, Y. J., R. Powers, and G. T. Montelione. 2005. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665-1674.
34. Huang, Y. J., R. Tejero, R. Powers, and G. T. Montelione. 2006. A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 62:587-603.
35. Jansson, M., Y. C. Li, L. Jendeberg, S. Anderson, B. T. Montelione, and B. Nilsson. 1996. High-level production of uniformly 15N- and 13C-enriched fusion proteins in Escherichia coli. J Biomol NMR 7:131-141.
36. Juge, N. 2012. Microbial adhesins to gastrointestinal mucus. Trends Microbiol 20:30-39.
37. Kamruzzaman Munshi, M., Y. Kobayashi, and T. Shikanai. 2005. Identification of a novel protein, CRR7, required for the stabilization of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. Plant J 44:1036-1044.
38. Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.
39. Kay, L. E., D. A. Torchia, and A. Bax. 1989. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28:8972-8979.
40. Li, N., E. A. Sickmier, R. Zhang, A. Joachimiak, and S. W. White. 2002. The MotA transcription factor from bacteriophage T4 contains a novel DNA-binding domain: the 'double wing' motif. Mol Microbiol 43:1079-1088.
41. Li, N., W. Zhang, S. W. White, and R. W. Kriwacki. 2001. Solution structure of the transcriptional activation domain of the bacteriophage T4 protein, MotA. Biochemistry 40:4293-4302.
42. Linge, J. P., M. A. Williams, C. A. Spronk, A. M. Bonvin, and M. Nilges. 2003. Refinement of protein structures in explicit solvent. Proteins 50:496-506.

43. Liu, J., G. T. Montelione, and B. Rost. 2007. Novel leverage of structural genomics. Nat Biotechnol 25:849-851.
44. Liu, Y., and J. H. Prestegard. 2010. A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins. J Biomol NMR 47:249-258.
45. MacKenzie, D. A., L. E. Tailford, A. M. Hemmings, and N. Juge. 2009. Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. J Biol Chem 284:32444-32453.
46. Moseley, H. N., D. Monleon, and G. T. Montelione. 2001. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol 339:91-108.
47. Muraki, N., J. Nomata, K. Ebata, T. Mizoguchi, T. Shiba, H. Tamiaki, G. Kurisu, and Y. Fujita. 2010. X-ray crystal structure of the light-independent protochlorophyllide reductase. Nature 465:110-114.
48. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536-540.
49. Neri, D., T. Szyperski, G. Otting, H. Senn, and K. Wüthrich. 1989. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 28:7510-7516.
50. Ogawa, T., and H. Mi. 2007. Cyanobacterial NADPH dehydrogenase complexes. Photosynth Res 93:69-77.
51. Otwinowski, Z., and W. Minor. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology 276:307-326.
52. Peng, L., W. Cai, and T. Shikanai. 2010. Chloroplast stromal proteins, CRR6 and CRR7, are required for assembly of the NAD(P)H dehydrogenase subcomplex A in Arabidopsis. Plant J 63:203-211.

53. Peng, L., and T. Shikanai. 2011. Supercomplex formation with photosystem I is required for the stabilization of the chloroplast NADH dehydrogenase-like complex in Arabidopsis. Plant Physiol 155:1629-1639.
54. Peng, L., H. Shimizu, and T. Shikanai. 2008. The chloroplast NAD(P)H dehydrogenase complex interacts with photosystem I in Arabidopsis. J Biol Chem 283:34873-34879.
55. Perrakis, A., R. Morris, and V. S. Lamzin. 1999. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6:458-463.
56. Rückert, M., and G. Otting. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. Journal of the American Chemical Society 122:7793-7797.
57. Sarma, R., B. M. Barney, T. L. Hamilton, A. Jones, L. C. Seefeldt, and J. W. Peters. 2008. Crystal structure of the L protein of Rhodobacter sphaeroides light-independent protochlorophyllide reductase with MgADP bound: a homologue of the nitrogenase Fe protein. Biochemistry 47:13004-13015.
58. Schrödinger, L., posting date. The PyMOL Molecular Graphics System, Version 0.99rc6. [Online.]
59. Schwieters, C., J. Kuszewski, and G. Clore. 2006. Using Xplor-NIH for NMR molecular structure determination. Progress in Nuclear Magnetic Resonance Spectroscopy 48:47-62.
60. Sheldrick, G. M. 2008. A short history of SHELX. Acta Crystallogr A 64:112-122.
61. Shen, Y., F. Delaglio, G. Cornilescu, and A. Bax. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213-223.
62. Singarapu, K. K., G. Liu, R. Xiao, C. Bertonati, B. Honig, G. T. Montelione, and T. Szyperski. 2007. NMR structure of protein yjbR from Escherichia coli reveals 'double-wing' DNA binding motif. Proteins 67:501-504.

63. Terwilliger, T. C. 2003. Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr D Biol Crystallogr 59:38-44.

64. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

65. Tjandra, N., S. Grzesiek, and A. Bax. 1996. Magnetic field dependence of nitrogen-proton J splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. Journal of the American Chemical Society 118:6264-6272.

Chapter 5: Conclusions

5.1 HetR- and PatS-dependent regulation of heterocyst differentiation in cyanobacteria

The regulatory networks controlling the initiation, development, and maintenance of cyanobacterial heterocysts represent some of the most archaic examples of cell differentiation found in nature. Cellular differentiation and morphogenesis are biological concepts commonly thought of as unique to the eukaryotic world, yet heterocyst cells of cyanobacteria are a great example of cellular differentiation at work in a primitive bacterium. In the last two decades, there has been a steady increase in the number of gene elements discovered to have potential roles in heterocyst development as described in recent reviews (5, 7, 8); however, there is still a deficiency of detailed molecular characterization of the regulatory components involved. The work presented in this dissertation describes our efforts to better understand the regulation of cyanobacterial heterocyst differentiation at the molecular level.

It is clear that HetR and PatS are important regulatory elements in cyanobacteria. The observance that both the *hetR* and *patS* genes are present in heterocystous and non-heterocystous cyanobacteria (9) has opened the door to new questions about their function across different species, and made the interpretation of their biological roles even more complex. We chose to take a biophysical approach to specifically focus on the structure-function relationship of HetR- and PatS-dependent regulation, of which we were able to address several outstanding fundamental questions: 1) Can HetR from *Anabaena* be expressed soluble and isolated in sufficient yield for structural studies while retaining its activity?, 2) What is the oligomeric state of HetR in solution?, 3) Does HetR change its oligomeric state when bound to its DNA substrate?, 4) What is the binding partner for the PatS inhibitor peptide?, and 5) What is the functional form of PatS in solution?

The pursuit of these questions led to the discovery of many important findings. In Chapter 2, we were able to provide a unique experimental method to show the first

unambiguous evidence that PatS binds directly to HetR, even in the absence of DNA (3). This binding interaction exists as one HetR homodimeric protein bound to two molecules of PatS peptide (3). We were also able to show evidence that HetR exists in solution as a stable homodimer and binds to one molecule of double stranded DNA as a dimer (3).

While testing our new hypotheses (Chapter 3), we were able to quantify the binding thermodynamics of various PatS C-terminal peptide fragments and showed that PatS-6 binds an order of magnitude tighter than the PatS-5 fragment (2). Our results suggest that the enhancement in binding affinity is likely driven by a backbone interaction with the PatS-6 N-terminal glutamate atoms and HetR (2). Conversely, the PatS-7 fragment bound three orders of magnitude weaker than PatS-6, and the PatS-8 peptide was incapable of binding to HetR (2).

It is important to understand the limitations of our experimental observations however. Despite the recent report of the X-ray crystal structure of HetR from the thermophile *Fischerella* MV11 (6), the HetR structure from *Anabaena* remains unknown. Structures of any HetR-substrate complexes are also unknown. Therefore, the specific locations of the MTSL spin labels we have modeled in the homologous structure for biophysical studies may not necessarily represent the actual sites and local topology for *Anabaena* HetR, described in Chapter 2. We also recognize that the influence of the MTSL spin label dynamics we observed following an introduction of PatS peptides may also be the result a distal allosteric effect in the protein upon PatS ligand docking, rather than a change influenced by direct contact with PatS in a putative binding site near the spin probe. In addition, our observations that the PatS-8 peptide failed to bind under the conditions tested, does not rule out the possibility that PatS sequences greater than or equal to PatS-8 may still bind to HetR under the right conditions *in vitro* or *in vivo*. The same could be true for any length of the PatS peptide to still potentially bind DNA and even form a ternary complex with HetR or other proteins under the right conditions.

These observations provide the foundation for future *in vitro* and *in vivo* experiments to identify the structures, binding sites, and binding mechanisms of HetR,

PatS, DNA, and any other regulatory partners involved in forming complexes, in addition to identifying the active form of HetR and PatS inside a cyanobacterial cell. This research is of particular importance to society based on the pivotal roles that cyanobacteria have played and continue to serve for life on planet Earth. The dual functions of photosynthesis and nitrogen fixation that many cyanobacteria serve have important potential impacts for the future from designing more efficient sources of energy to utilization as effective crop fertilizers. Furthermore, the regulatory networks that control cell differentiation serve as excellent models for higher eukaryotic signaling pathways which have critical roles in cell biology, and in turn carry high impacts in life sciences research, from embryonic development to therapeutics.

5.2 Structural Genomics

The advent of the Protein Structure Initiative and subsequent appearance of high-throughput structural genomics consortia has led to a wealthy contribution of information to the sciences in the last 12 years. The overarching goal of the PSI, to make available the three-dimensional structure of most proteins based on their gene sequences, is a daunting task but significant progress has been made to accomplish this tremendous feat (1). The work presented and discussed in this dissertation is a small but integral piece to the vast puzzle encompassing the central dogma of molecular biology. The worldwide cooperation of structural genomics groups, like the Northeast Structural Genomics consortium, has expedited the answers to many questions in biology that require a detailed structural understanding of biological macromolecules. Naturally, as has always been the case with scientific research, the answers to many such protein structural biology postulations, reveals several new questions from each respective structure solved.

The four studies discussed here in Chapter 4 contribute to the goals set forth by the National Institute of General Medical Sciences. Our efforts to solve high resolution solution NMR and X-ray crystal structures of proteins Pspto_3016 from the plant

pathogen *Pseudomonas syringae*, the cyanobacterial proteins, Asl3597 (4) and Asr4154 from *Anabaena* sp. PCC 7120, and lastly the mucin-binding domain of protein LBA1460 from *Lactobacillus acidophilus*, have provided the foundation for future biological and biochemical characterization for each of these different gene targets. The functional analysis of each of these uncharacterized proteins has been studied from a bioinformatics perspective, combined with the information garnered from high resolution three-dimensional structures of proteins both solution and in the crystalline state. We are thus, of course, limited as to how much functional information can be determined from the data available. To make these 3D structures available to the public, especially for those from newly identified and uncharacterized domains of unknown function will assist in future functional studies to come.

5.3 References

1. Burley, S. K., A. Joachimiak, G. T. Montelione, and I. A. Wilson. 2008. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. Structure 16:5-11.
2. Feldmann, E. A., S. Ni, I. D. Sahu, C. H. Mishler, J. D. Levengood, Y. Kushnir, R. M. McCarrick, G. A. Lorigan, B. S. Tolbert, S. M. Callahan, and M. A. Kennedy. 2012. Differential binding between PatS C-terminal peptide fragments and HetR from Anabaena sp. PCC 7120. Biochemistry 51:2436-2442.
3. Feldmann, E. A., S. Ni, I. D. Sahu, C. H. Mishler, D. D. Risser, J. L. Murakami, S. K. Tom, R. M. McCarrick, G. A. Lorigan, B. S. Tolbert, S. M. Callahan, and M. A. Kennedy. 2011. Evidence for Direct Binding between HetR from Anabaena sp. PCC 7120 and PatS-5. Biochemistry 50:9212-9224.
4. Feldmann, E. A., T. A. Ramelot, Y. Yang, R. Xiao, T. B. Acton, J. K. Everett, G. T. Montelione, and M. A. Kennedy. 2011. Solution NMR structure of Asl3597 from Nostoc sp. PCC7120, the first structure from protein domain family PF12095, reveals a novel fold. Proteins.

5. Golden, J. W., and H. S. Yoon. 2003. Heterocyst development in Anabaena. Curr Opin Microbiol 6:557-563.
6. Kim, Y., G. Joachimiak, Z. Ye, T. A. Binkowski, R. Zhang, P. Gornicki, S. M. Callahan, W. R. Hess, R. Haselkorn, and A. Joachimiak. 2011. Structure of transcription factor HetR required for heterocyst differentiation in cyanobacteria. Proc Natl Acad Sci U S A 108:10109-10114.
7. Kumar, K., R. A. Mella-Herrera, and J. W. Golden. 2010. Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol 2:a000315.
8. Zhang, C. C., S. Laurent, S. Sakr, L. Peng, and S. Bédu. 2006. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. Mol Microbiol 59:367-375.
9. Zhang, J. Y., W. L. Chen, and C. C. Zhang. 2009. hetR and patS, two genes necessary for heterocyst pattern formation, are widespread in filamentous nonheterocyst-forming cyanobacteria. Microbiology 155:1418-1426.