Title
Structural and biochemical studies of sigma54 transcriptional activation in Aquifex aeolicus

Author
Vidangos, Natasha Keith

Publication Date
2010

Peer reviewed|Thesis/dissertation

Structural and biochemical studies of σ^{54} transcriptional activation in *Aquifex aeolicus*

by
Natasha Keith Vidangos

A dissertation in partial satisfaction of the degree of
Doctor of Philosophy
in
Chemistry

in the
Graduate Division
of the University of California, Berkeley

Committee in charge:
Professor David Wemmer, Chair
Professor Jennifer Doudna
Professor Sydney Kustu

Fall 2010

Abstract

Structural and biochemical studies of σ^{54} transcriptional activation in *Aquifex aeolicus*
by
Natasha Keith Vidangos
Doctor of Philosophy in Chemistry
University of California, Berkeley
Professor David E. Wemmer, Chair

This thesis addresses a diversity of questions regarding the structural details of σ^{54} transcriptional activation, and the function of σ^{54} activation in the hyperthermophile *Aquifex aeolicus*. In order to place each topic in its appropriate context, a general introduction is provided in the first chapter, and supplemented with additional, more detailed introductions in each subsequent chapter. The second chapter reflects the central project of this thesis, the determination of the structure of the DNA-binding domain of an NtrC-like σ^{54} transcriptional activator protein in *Aquifex aeolicus*, in complex with its high-affinity DNA binding site. Although this project was attempted by NMR, its structure was ultimately solved by X-ray crystallography. This structure, which shows slight DNA-bending, is compared to the recent structure of the homologous Fis protein in complex with DNA.

In the third chapter, I describe a cross-comparison of DNA-binding domains from *Aquifex aeolicus*, including new structures of the DNA-binding domains of NtrC1 and NtrC2, and comparisons to NtrC4, ZraR and NtrC from *Salmonella enterica serovar typhimurium*, and Fis from *Escherichia coli*. These structures of DNA-binding domains from a single family enable a detailed comparison of structural changes that tune protein function. A trend is noted, in which larger dimerization interfaces in the DNA-binding domains appear to correlate with smaller dimerization interfaces in the other domains of the protein. The additional structure of the full-length NtrC1 protein is presented, but the DNA-binding domains were missing from the density, providing evidence for the flexibility of the linker between the central and DNA-binding domains. Together with structural information about the CD linker regions at the N-termini of the DNA-binding domains, I conclude that the CD linker regions are generally unstructured in the inactive state, functioning as tethers to bring the RC domains of NtrC into appropriate local concentrations for activity.

In the fourth chapter, I depart from σ^{54} transcriptional activators, and discuss NMR studies on the σ^{54} factor. The Darst lab at Rockefeller University has produced a crystal structure of the full-length intact σ^{54} factor in complex with DNA. However, poor data quality prevents regions of the molecule from being traced unambiguously through the density. By applying new methods in NMR including TROSY spectroscopy and specific isotopic labeling, I attempted to resolve small regions of structure in this ambiguous region. However, these studies were complicated by protein aggregation.

The fifth and final chapter takes a step back from structural perspective and compiles and discusses our current understanding of σ^{54} regulatory networks in *Aquifex aeolicus*. This analysis of genes with bioinformatics and biochemical techniques led me

to discover that NtrC3, an unstudied σ^{54} transcriptional activator, binds upstream of the dhsU gene, which is responsible for sulfur metabolism. NtrC3 also appears to be associated with a heme-binding, soluble histidine kinase, HksP4. In order to explore this two-component regulatory system further, the HksP4 protein was prepared and its gas-binding properties were studied in collaboration with Dr. Brian Smith of the Marletta lab. We find that the heme binds O_2, NO, and CO gases in the reduced ferrous iron state, but did not observe any autophosphorylation activity in any state. I conclude that an additional element, such as an additional activator, or more physiological conditions, would be necessary for activity.

Approved:

Chair: _____ Date: _____

To my grandfather, Serge Zaroodny, the Russian physicist I never met,
but for whom a doctorate degree meant everything in the world.

Acknowledgements

As with any thesis, there are so many people to thank. I thank my brothers for sharing in the stories, and my parents for their love and encouragement in the sciences for all these years, and their support and insistence that I should "hang on and finish". I thank my husband, Ivan, for supporting me with good heart and humor through the six long years of a cross-costal long distance relationship, and with his insistence that I could "quit at any moment" and everything would still be all right. I thank my advisor David Wemmer, whose hands-off style, infinite patience, scientific rigor, and generous flexibility were deeply influential factors in my learning and enjoyment of my degree. Perhaps it is informative to say that in appreciation for Dave's skill as a teacher and mentor, our lab nominated Dave for a graduate student mentoring award not once, but twice. I want to thank my old labmates for making the experience so fulfilling and the working environment harmonious. I thank Aaron Philips for his good humor and spirit that brought the Wemmer lab together with laughter, music, and West Virginian slang. I thank Jeff Pelton for hours of help with every aspect NMR. I thank Joe Batchelor for years of advice and shortcuts, and for the myriad of lessons I learned from him about letting go of old mistakes, learning to dream up new scientific ideas, and having a healthy supply of chocolate on hand for any emergencies.

Then there was the series of strong, inspiring women who stepped in and became unasked-for, unlooked-for mentors when I needed them the most. In my first years in graduate school, it was Michaeleen Callahan-Doucleff, my "big-sister" graduate student who taught me everything I know and knew when to send me off into the woods, and when to sit me down with a plan. I took great comfort from my many professional soul-searching conversations and milk teas with my labmate Judy Hwang. And then there was Dr. Sonja Lorenz, who shared her enthusiasm with me in Schumann and Mendelssohn but also gave me unlimited scientific and personal wisdom on a daily basis, as well as proofreading a large part of this thesis. I have an especially deep gratitude to Professor Sydney Kustu, one of the most inspiring people I have known, who discovered me at the pit of a research slump, and taught me how to "stand up" and remember to be passionate about my work. I'm also grateful to Terry Lang for her random acts of kindness, Monica Smith for years of selfless support, and Terry Yen for endless friendship, laughs, and teaching me the mantra, "don't stress – organize!"

Scientifically, many people had a hand in the work presented in this thesis. Thanks goes to Dr. Brian Smith of the Marletta lab for hours of advising on the behavior of heme proteins. Thanks also to Dr. Artem Lyubimov for his humor, rigor, and generosity with his time in working with me on X-ray crystallography – even joining me on my graveyard shifts at the beamline. I thank the Kuriyan, Marletta, and Marqusee labs for untold amounts of shared chemicals, shakers, media tubes, crystallography materials, instrument time, and endless amounts of advice. I thank the two spectacular undergraduates, Jimmy Ton and Ahmed Alkholeidi, who worked for me through thick and thin, and contributed a great deal to the final chapter of this thesis. I also thank the newer members of the lab: Alex, Zhijuan, Aleks, Hagit, and Wenshu, for making lab feel more and more like family. I'd also like to thank the members of the Science, Technology, and Engineering Policy (STEP) group for giving me the courage to explore a new professional dream.

Table of Contents

Chapter 1: Introduction to sigma54-dependent transcriptional regulation 1
- 1.1 Transcription 1
- 1.2 The σ^{70} class of σ factors 1
- 1.3 The σ^{54} factor 4
- 1.4 Structural understanding of σ^{54} 4
- 1.5 σ^{54}-dependent transcriptional activators 5
- 1.6 Structure and function of σ^{54} transcriptional activators 7
- 1.7 Overview of two-component signal transduction. 8
- 1.8 Thesis outline 9

Chapter 2: DNA-recognition by a σ^{54} transcriptional activator from *Aquifex aeolicus* 10
- **2.1 Summary** 10
- **2.2 Introduction** 10
  - 2.2.1. Structures of σ54-activator DNA-binding domains 10
  - 2.2.2. NtrC and Fis are homologues 10
- **2.3 Materials and Methods** 11
  - 2.3.1 Protein expression 11
  - 2.3.2. Protein Purification 12
  - 2.3.3. NMR of 1DBD and 4DBD complexes 12
  - 2.3.4. Complex preparation for crystallography 13
  - 2.3.5. Crystallography of 4DBD-complex 13
  - 2.3.6. Binding affinity assays 14
- **2.4 Results** 14
  - 2.4.1. NMR of 1DBD and 4DBD complexes 14
  - 2.4.2. Crystal structure of free-4DBD 15
  - 2.4.3. Determination of the DNA-binding site for NtrC4 19
  - 2.4.4. Structure determination of the 4DBD-site_1 complex 19
  - 2.4.5. 4DBD-complex: protein-DNA interactions 19
  - 2.4.6. 4DBD-complex: DNA conformation overview 21
  - 2.4.7. Biochemical analysis of DNA-binding site of NtrC4 22
- **2.5. Discussion** 22
  - 2.5.1. The relevance of lpxC activation to NtrC4 activity 22
  - 2.5.2. Comparison of 4DBD free protein to Fis 22
  - 2.5.3. Comparison of the 4DBD and Fis complex structures: protein interactions 23
  - 2.5.4. Comparison of the 4DBD and Fis complex structures: DNA distortion 26
  - 2.5.5. Conclusions on the mode of NtrC4 DNA-binding/bending 27

Chapter 3: NtrC-like C-terminal DNA-binding domains: structure, function, and flexible tethering to the central domain 29
3.1 Summary 29
3.2 Introduction 29
3.2.1 DNA-binding domains of σ54 activators 29
3.2.2 Functions of the DNA-binding domain 29
3.2.3 CD-linker regions 30
3.3 Materials and Methods 32
3.3.1 Constructs 32
3.3.2 Protein Expression and Purification 33
3.3.3 Analytical ultracentrifugation 33
3.3.4 Structural comparisons 33
3.4. Results 33
3.4.1 Construct selection of 1DBD and 2DBD 33
3.4.2 The structures of NtrC1-DBD and NtrC2-DBD 33
3.4.3 Dimerization interfaces: 1DBD and 2DBD are monomeric in solution 36
3.4.4 Crystal structure of full-length NtrC1: the DBD is flexibly tethered to the C domain 36
3.4.5 Cooperativity of binding in NtrC4-FL. 37
3.5 Discussion 39
3.5.1 DBD and RC dimer stability complement one another 39
3.5.2 DNA-binding and bending 42
3.5.3 Absence of prolines in *A. aeolicus* NtrCs 42
3.5.4 NtrC DBDs and cooperative binding 42
3.5.5 CD linkers of σ^{54} activators are disordered 43
3.6 Future studies 43
3.6.1 NMR studies on the N-terminal linker of 1DBD-long 43

Chapter 4: NMR studies on region $\sigma^{N}3$ of σ^{54} 44
4.1 Summary 44
4.2 Introduction 44
4.2.1 Structural studies on the $\sigma^{N}3$ region of the σ^{54} factor: current understanding 42
4.2.2 X-ray crystallography of the σ^{54} factor 42
4.2.3 Introduction to NMR 46
4.2.3.1 NMR Basics 46
4.2.3.2 T1, T2 relaxation processes 46
4.2.3.3 Relaxation mechanisms: dipolar relaxation and chemical shift anisotropy (CSA) 47
4.2.3.4 TROSY-NMR 48
4.2.3.5 Specific labeling strategies 49
4.3 Methods 52
4.3.1 Expression and purification of isotopically-labeled protein samples 52

4.3.2 NMR experiments 53
4.4 Results 53
4.4.1 Optimization of production strategies of isotopically-labeled samples 53
4.4.2 NMR on 69-398 constructs 55
4.5 Discussion 64
4.5.1 Spectral quality and protein aggregation 64
4.5.2 Future studies 65

Chapter 5: Exploring two-component regulatory networks in *Aquifex aeolicus* 66
5.1 Summary 66
Part 1: Gene-association studies in *Aquifex aeolicus* 66
5.1.1 Introduction 66
5.1.1.1 Introduction to function-correlation studies in *A. aeolicus* 66
5.1.1.2 Determination of gene associations and putative binding sites 67
5.1.2. Materials and Methods 68
5.1.2.1 Protein expression and purification 68
5.1.2.2 Consensus sequence searches 69
5.1.2.3 PCR, cloning of binding sites into Pet21a vector 69
5.1.2.4 Native protein-DNA gel-shift assays 69
5.1.3 Results and Discussion 69
5.1.3.1 Search for genes under σ^{54} control in *A. aeolicus* 69
5.1.3.2 Functional roles of *A. aeolicus* σ^{54}-regulated genes 71
5.1.3.3 Gene-by-gene analysis 71
5.1.3.4 Determination of the DNA-binding sites for NtrC3 and NtrC4 76
5.1.3.5 Implications of NtrC3, NtrC4, NifA regulation 78
5.1.3.6 The absence of additional binding associations 78
5.1.3.7 Current model for σ^{54} gene associations in *A. aeolicus* and future studies 79
5.1.3.8. Distances between activator and σ^{54} binding sites 79
Part 2: Characterization of a histidine kinase: HksP4 81
5.2.1 Introduction 81
5.2.1.1 Histidine kinases in TCS pathways: common structural/functional themes 81
5.2.1.2 Heme-containing histidine kinases 81
5.2.1.3 Introduction to heme biochemistry 85
5.2.1.4 Characterization of a histidine kinase in *A. aeolicus*: HksP4 85
5.2.2. Materials and Methods 87
5.2.2.1 Cloning 87
5.2.2.2 Protein expression of HksP4-FL and HksP4-PAS 87
5.2.2.3 Activity assays on HksP4 87
5.2.3. Results and Discussion 89

5.2.3.1 HksP4 contains heme, and binds O_2, NO, and CO, but not H_2S 89
5.2.3.2 Activity assays: anaerobic, O_2-, NO-, and CO-bound HksP4 do not exhibit ATPase activity 91
5.2.4. Concluding remarks 91

References 92
Appendices 106
Appendix A: Viscosity of glycerol:water mixtures at varying temperatures 106
Appendix B: Communication from Professor Valley Stewart, UC Davis, on the function of the nirB and fhp gene cluster 107

Chapter 1: Introduction to σ^{54}-dependent transcriptional regulation

1.1 Transcription

The ability to adapt to environmental changes is a necessary skill for the survival of any organism. This is particularly true of single-celled organisms, which are not able to buffer themselves against external changes with additional protective cells or tissues. In prokaryotes, the regulation of gene expression to respond to external stimuli includes transcriptional, translational, and post-translational mechanisms. However, transcription is the most dynamically regulated point of control.

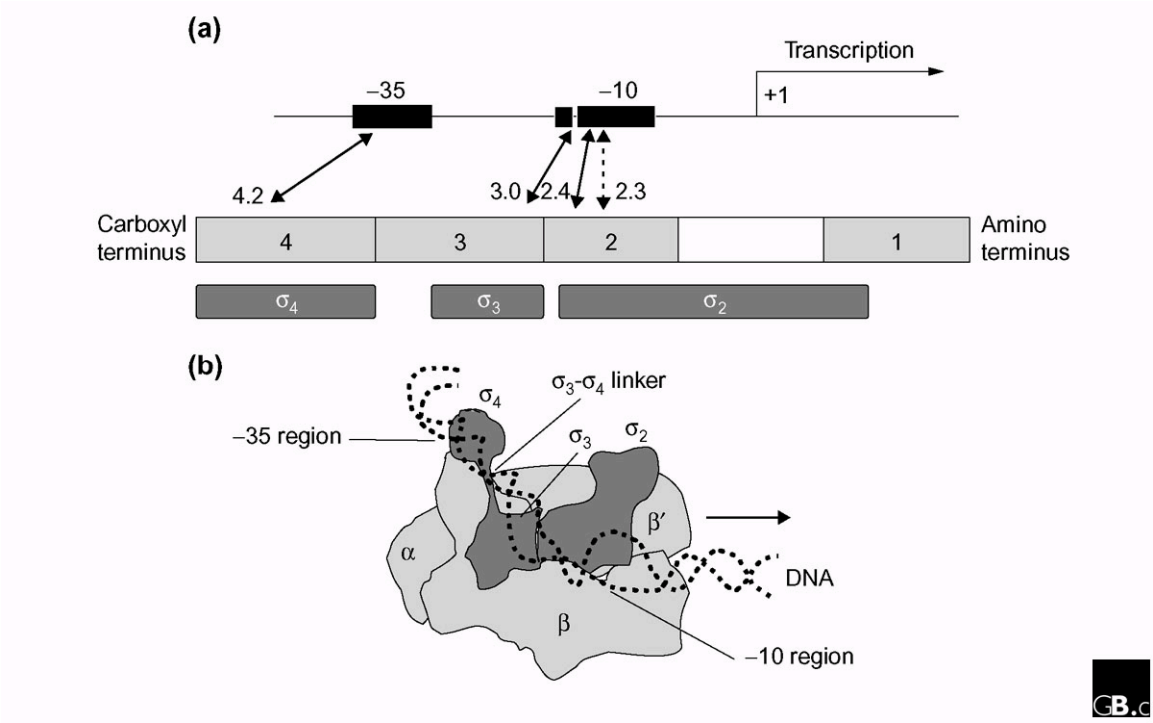
Transcription requires the activity of RNA Polymerase (RNAP), a multisubunit ~400kDa protein. RNAP consists of 5 core subunits, with the architecture $\alpha_2\beta\beta'\omega$. This core enzyme is capable of transcriptional elongation and termination of mRNA constructs, but cannot bind to promoter DNA specifically and therefore cannot initiate transcription (Buck, et al., 2000; Gosh, et al., 2010). To form the active holoenzyme the core RNAP reversibly binds to a sixth subunit called a σ factor. The σ factor conveys DNA-binding specificity, positioning the RNAP-holoenzyme in the promoter region of the gene being transcribed. A variety of σ factors exist in any given organism, and can be regulated by different mechanisms. This enables a cell to tune the transcriptional regulation of different gene classes.

The activity of σ factors can be controlled by a number of different processes on their production: transcriptional and translational regulation, post-translational processing, and proteolytic degradation (Paget, 2003). Another method of repression of σ factors is the post-translational inhibition by anti-σ factors, proteins that bind the σ factor and prevent it from associating with the core RNAP (Campbell, 2008).

1.2 The σ^{70} class of σ factors

There are two structurally and functionally distinct classes of σ factors, σ^{70} and σ^{54}. The σ^{70} class is the most common, and includes a wide diversity of subclasses. A number of σ^{70}-class factors are often found in a single genome, with as many as 63 found in the genome of *Streptomyces coelicolor* (Paget, 2003). The σ^{70} class can be divided into four groups. Group 1 factors, also known as primary factors, are responsible for vital processes such as housekeeping genes, metabolism, and development. Group 2 factors are dispensable for cell growth, and regulate processes such as ancillary metabolism. Group 3 factors are usually used to respond to a specific signal to perform stress-response functions such as sporulation or heat-shock response. Group 4 includes the diverse extracytoplasmic function (ECF) subfamily, which regulate signals from the extracytoplasmic environment (Paget, 2003). A number of σ^{70} factors have been implicated in bacterial virulence, including σ^{B} from *Bacillus anthracis* and *M. tuberculosis*, RpoE from *H. influenzae*, *Salmonella enterica serovar Typhimurium*, and *Vibrio cholerae*, and the factors $\sigma^{C}/\sigma^{D}/\sigma^{E}/\sigma^{F}$ from *M. tuberculosis* (Kazmierczak, 2005).

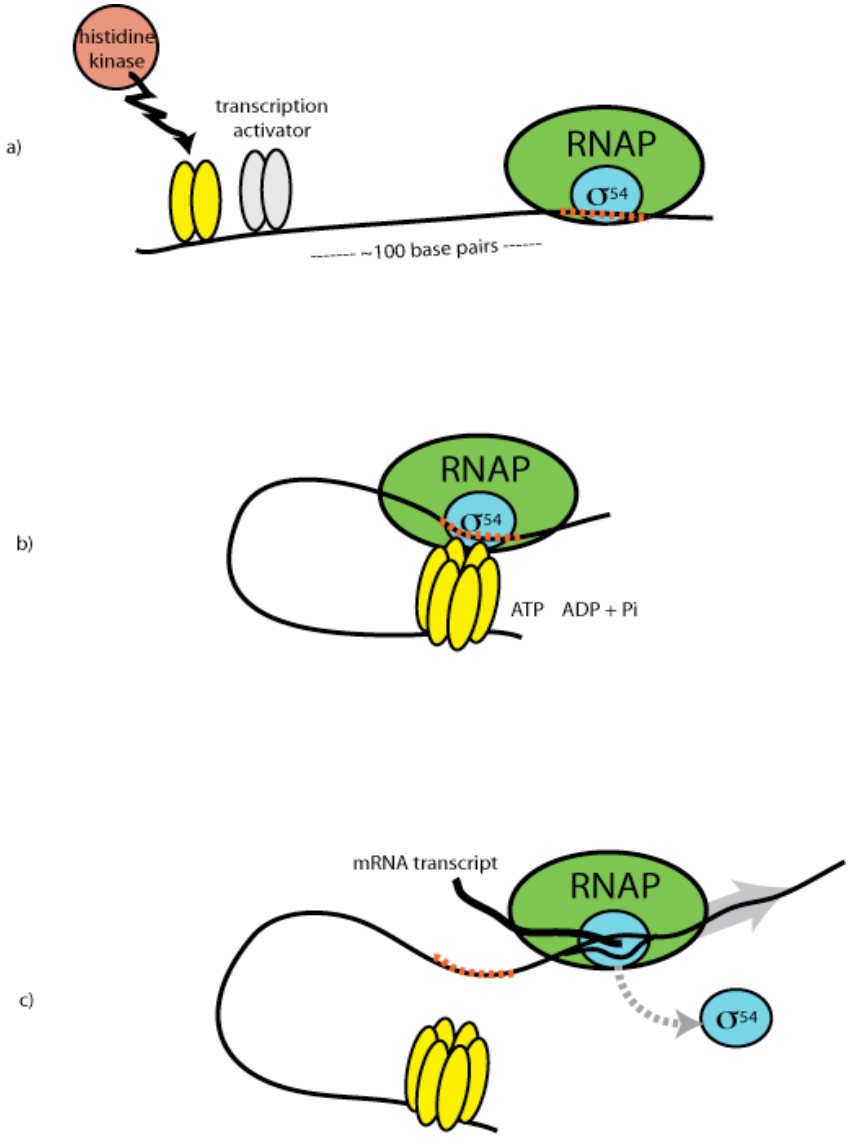
Figure 1.1. Domain architecture of σ^{70} and schematic of σ^{70}-RNAP-holoenzyme. a) the architecture of a σ^{70} promoter region, and the σ^{70} domains that interact with each promoter element. Two nomenclatures for domain regions are used: the σ1-σ4 regions are discussed in the text. b) Schematic of the σ^{70}-RNAP holoenzyme open complex. Reproduced from Paget, 2003.



The σ^{70} factor binds to two conserved promoter DNA regions approximately –10 and –35 base pairs upstream of the transcription initiation site at a consensus sequence TTGACA-$N_{15\text{-}19}$-TATAAT (Lisser and Margalit, 1993). σ^{70} contains four flexibly-linked structural regions, called σ1, σ2, σ3, and σ4 (figure 1.1). Region σ1 is conserved only between closely related σ^{70} factors, and appears to serve to antagonize the DNA-binding activity of the σ factor (Paget, 2003). The σ2 and σ3 regions cooperate in the recognition of the –10 binding region, and a subregion of σ2 (called 2.3) is responsible for promoter melting (Barne, et al., 1997). The C-terminal end of region σ4 contacts the –35 binding region. Structures have been solved of two primary σ^{70} factors and one ECF σ factor: σ^{70} from *E. coli* (Malhotra, et al., 1996), σ^{A} from *Thermus Aquaticus* (Campbell, *et al.*, 2002), and σ^{R} from *S. coelicolor* (Wei, et al., 2002). With the exception of the σ1 domain, which is poorly conserved between these organisms, and the absence of region σ3 from the ECF domain, the structures were very similar. A recent crystal structure of the σ^{70} factor in the holoenzyme complex shows that the σ2, σ3, and σ4 regions all provide contacts with core RNAP, creating a buried surface region of over 8000 $Å^2$.

The current understanding of the σ^{70} mechanism is as follows. The σ^{70} factor first associates with the RNAP core enzyme, then binds to the –10 and –35 regions upstream of the transcription initiation site via regions σ2 and σ4. As a result, the DNA in the

Figure 1.2. Schematic for the mechanism of σ^{54} transcriptional activation. a) An inactive transcriptional activator binds 100 base pairs upstream of the σ^{54}-RNAP-holoenzyme, and is activated, often by phosphorylation by a histidine kinase. b) Activator activation, ATP hydrolysis, and remodeling of σ^{54}-RNAP-holoenzyme. c) Transcriptional initiation.



region –11 to +4, which includes a portion of the –10 interaction site, is melted. Once melted, the template and non-template strands are separated: the template strand interacts with the RNAP active site, and the non-template strand passes through a number of stabilizing, aromatic residues in the σ2-2.3 region. After a nascent RNA of 8-10 nucleotides is synthesized, the σ factor is released, and RNAP escapes the promoter (also known as promoter clearance) to begin elongation. The disordered loop between regions σ3 and σ4 may play a role in stabilizing the open complex, and/or the mechanism of promoter clearance (Paget, 2003; Ghosh, et al., 2010).

1.3 The σ^{54} factor

The σ^{54} factor was first discovered as a regulator for nitrogen assimilation in enteric bacteria (Hunt and Magasanik, 1985, Hirschmann et al., 1985). Since then, the σ^{54} factor has been identified in a number of distantly related organisms, ranging from *Bacillus subtilis*, *Chlamydia spp.,* and *Borrelia burgdorferi*, to *Aquifex aeolicus* (Studholme and Buck, 2000). σ^{54} factors have been implicated in a number of diverse functions, including nitrogen assimilation and regulation, carbon source utilization, certain fermentation pathways, flagellar synthesis, and bacterial virulence (Kustu, et al., 1989, Kazmierczak et al., 2005). Compared to the σ^{70} factor, the σ^{54} factor is relatively uncommon: bacteria encode an average of only 1 σ^{54} factor per genome (Buck, et al., 2000). For example, the *E. coli genome* encodes for 7 σ factors, including 6 σ^{70}-class factors, and 1 σ^{54} (Reitzer and Schneider, 2001).

The σ^{54} factor shares no detectable sequence homology with the σ^{70} factor, and it has no subclasses. It binds to a consensus sequence mrNrYTGGCACG-N4-TTGCWNNw (m=A/C, r=A/G, w=A/T, y=T/C, n=A/C/T/G; capitals indicate a high level of conservation, lowercase indicates moderate conservation), with conserved interactions occurring –12 and –24 base pairs upstream of the transcription initiation site (Barrios et al., 1999).

Unlike the σ^{70} factor, which is competent to initiate transcription when the σ^{70}-RNAP-holoenzyme is formed, the σ^{54}-RNAP-holoenzyme alone is unable to activate transcription. This is due to the fact that binding of σ^{54}-RNAP to DNA inhibits the melting of DNA. To enable DNA-melting, σ^{54} must interact with an ATP-dependent activator protein (section 1.5).

1.4 Structural understanding of σ^{54}

Like σ^{70}, the σ^{54} factor, which is also known as σ^{N}, consists of a number of functionally distinct, flexibly-linked domains (Wong, et al., 1994). Based on early studies of σ^{54} function, these domains were called Regions I (1-56), II (57-107), and III (108-477) (Buck, et al., 2000). In light of a decade of new data on σ^{54}, this nomenclature is outdated. Hence, this thesis will use a new nomenclature , analogous to σ^{70} domains, and refer to four regions called $\sigma^{N}1$ (*A. aeolicus* residues 1-69), $\sigma^{N}2$ (69-198), $\sigma^{N}3$ (198-300), and $\sigma^{N}4$ (300-398) (Figure 1.3). $\sigma^{N}1$ is implicated in interacting with the activator

protein, but little structural information is currently available for this region. σ^N2, also known as the core-binding domain, provides contacts to RNAP. Its structure was solved in our lab by solution NMR, and consists of a stacked three-helix and four-helix bundle, connected by a hinge point. This hinge region may be implicated in the RNAP-holoenzmye transition from the transcriptionally inhibited to transcriptionally active state (Hong, 2008). The structure of σ^N3 is not well understood, but is believed to serve functions similar to those of $\sigma2$ and $\sigma3$ in σ^{70}, including the interaction with the –12 region of DNA, and DNA melting (Wong, et al., 1994). σ^N4, like the $\sigma4$ region of σ^{70}, is a helix-turn-helix DNA-binding domain with specificity for the –24 region. Its structure, both unbound and bound to its target DNA-binding site, were solved in our lab by solution NMR (Doucleff et al., 2005; Doucleff et al., 2006). Although σ^{54} shares no detectable sequence homology with σ^{70}, the σ^N4 region in complex with DNA occupies a similar position on the DNA as the $\sigma4$ region of σ^{70}.

A number of cryo-EM structures have also been determined, including σ^{54} in complex with σ^{54}-activator PspF (Rappas et al., 2005), the σ^{54}-RNAP holoenzyme, and the σ^{54}-RNAP-PspF complex (Bose, et al., 2008). In their reconstructions of the σ^{54}-holoenzyme both free and in complex with PspF, Bose et al., identify four regions of σ^{54} density: D1, D2, D3, and Db (figure 1.4). On the basis of deletion mutants and nanogold labeling techniques, they assign D1 to region σ^N2, D3 to region σ^N4, and Db to region σ^N1. D2 is assigned to a combination of σ^N2 and σ^N3 regions. In the reconstruction of the free holoenzyme, region σ^N1 occupies a space between regions D1, D2, and D3. Docking DNA into the structure based on its position in σ^{70}-holoenzyme reconstructions shows that σ^N1 might sterically block the access of DNA to the RNAP active site. Upon binding to PspF, the σ^N1 density undergoes a dramatic shift to contact the activator protein. This is expected to bring DNA in contact with the RNAP active site, and enable transcription to initiate. Atomic-level details, however, including the role of σ^N3 and the exact nature of the structural rearrangement, are still poorly understood.

1.5 σ^{54}-dependent transcriptional activators

As described above, σ^{54} activation requires the presence of a transcriptional activator. One prototypical σ^{54}-activator is Nitrogen Regulatory Protein C (NtrC), which was first discovered to activate the transcription of glutamine synthetase genes in *E. coli* in conditions of nitrogen starvation (Hirschman, et al., 1985; Hunt and Magasanik, 1985). NtrC-like activators have three domains: a regulatory domain (R) that participates in two-component signal transduction, a highly conserved central AAA+ ATPase domain (C), and a dimeric helix-turn-helix DNA-binding domain (DBD). Transcriptional activators exist as dimers in the inactive state, and bind, often cooperatively, to two tandem DNA-binding sites located approximately 100 base pairs upstream of the TIS (Wedel et al., 1990). When the activator is activated, usually via phosphorylation of the regulatory domain by a histidine kinase, this stimulates the oligomerization of the full-length activator into a higher-order oligomer, usually a hexamer (Tucker and Sallai, 2007). The oligomer contacts the σ^{54} factor by looping out the intervening DNA, sometimes with the assistance of a DNA-bending protein such as Integration Host Factor (IHF) (Santero, et al., 1992). Upon contacting the σ^{54}-factor, the activator protein hydrolyzes ATP, and

Figure 1.3. a) Domain architecture and nomenclature used in the thesis. Solved structures of the σ^N2 and σ^N4 domains are shown below (Hong, et al., 2008; Doucleff, et al., 2007).

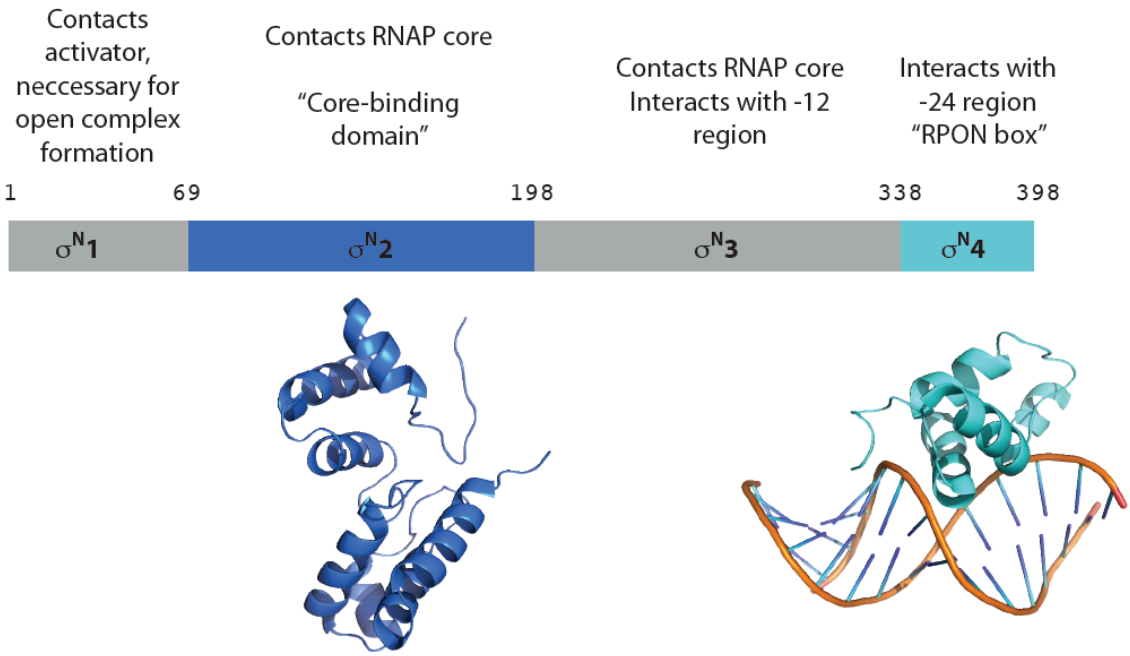
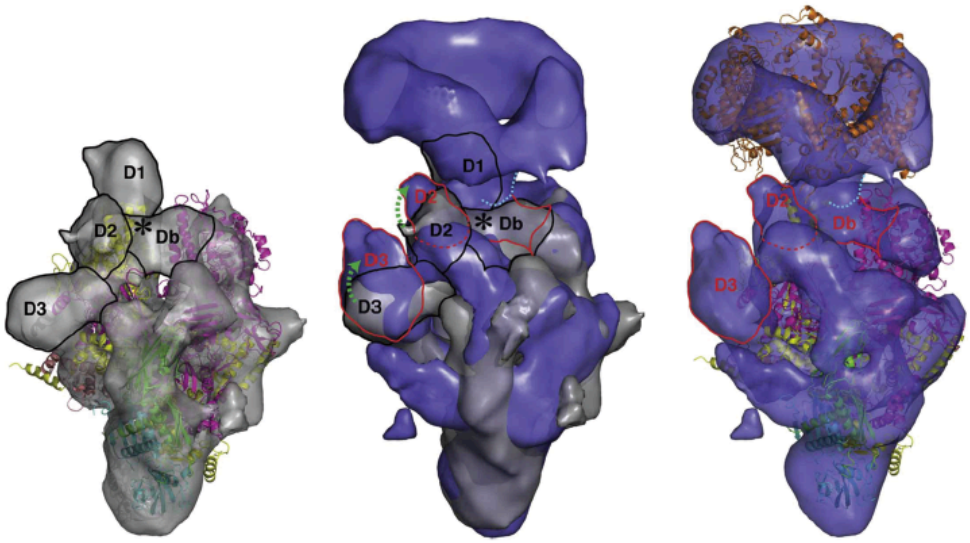


Figure 1.4. Cryo-EM density and modeling of the σ^{54} factor in complex with activator PspF. From Bose, et al., 2008.



remodels the σ^{54} factor (Wedel and Kustu, 1995). This structural change removes the inhibition of DNA melting, and enables transcriptional activation to occur. A schematic of this mechanism is shown in Figure 1.2.

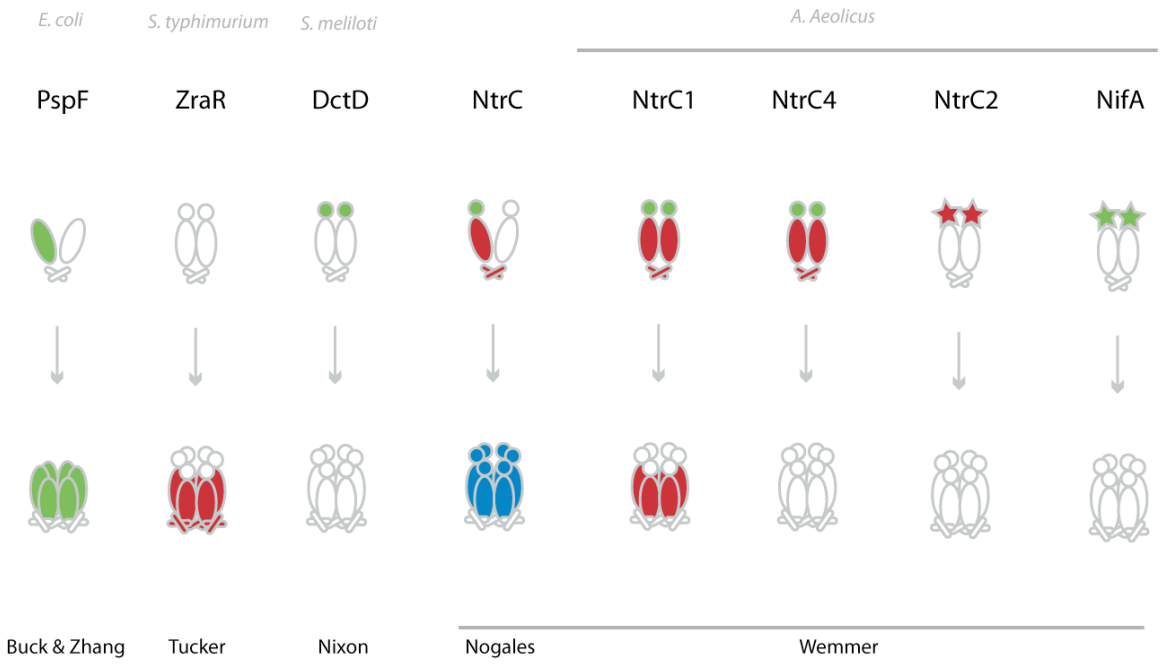
Although the NtrC-like architecture of Receiver-Central-DBD domains is common, the only domain that is required for σ^{54} activation is the highly-conserved central AAA+ ATPase domain. Variation at the regulatory and DNA-binding domains enables different σ^{54} activators to be activated differently, and bind DNA differently, to serve the regulatory requirements of a particular gene. Distinct regulatory domains have been observed in other σ^{54} activators, including GAF, PAS, CBS, V4R, and PRD domains. Also, multiple regulatory domains may be used, or the regulatory domain may be absent altogether (Studholme et al., 2003). The DNA-binding behavior of trancriptional activators also varies. Although the prototypical σ^{54}-activator protein binds to two tandem cooperative binding sites in the UAS, cases have been reported in which three or more binding sites are used to modulate activity. This includes the case of NtrC, in which additional binding sites in the UAS, called "governor sites", can be used to inhibit activity at the promoter, most likely by restricting DNA looping at high NtrC concentrations (Atkinson, et al., 2002). In the case of NorF, three activator binding sites are required for full activity (Tucker et al., 2010). Some activators, including FlgR from *Helicobacter pylori*, lack the DNA-binding domain altogether, and activate σ^{54} directly from solution (Brahmachary et al, 2004).

The mechanistic requirement for activator proteins signifies that σ^{54} provides a more tightly controlled regulatory mechanism than σ^{70}. Whereas σ^{70} activity is regulated by the concentrations of σ^{70} factors and anti-sigma factors within a restricted range, σ^{54} activity is predominantly modulated via the protein activators. This endows σ^{54} activation with a wide dynamic range. As a result, σ^{54}-dependent gene expression is often involved in immediate responses to environmental changes, such as chemotaxis, virulence factors, and alternative metabolism in starvation conditions (Kazmierczak, et al., 2005). The observation that this mechanism is highly conserved in a number of distantly-related bacteria underscores its evolutionary importance.

1.6 Structure and function of σ^{54} transcriptional activators

Currently available structures of σ^{54}-activators are summarized in figure 1.5. Considerable work on the regulatory domains (receiver and GAF) has been performed, resulting in the activated and unactivated solution NMR structures of DctD (Park et al., 2002), NtrC (Kern, *et al.,* 1997; Hastings, et al., 2003), NtrC1 (Lee, et al., 2002; Doucleff et al., 2005), NtrC4 (Batchelor, et al., 2008). Our lab has recently solved the GAF domain structures of NifA and NtrC2. Structures of the central domains of σ^{54}-activators have also been presented, including the heptameric central domain of NtrC1, (Lee, et al., 2003), and the RC unactivated dimers of NtrC1 and NtrC4. A structure of the oligomerized central and DNA-binding domains of ZraRwas also solved, although the DNA-binding domains form crystallographic dimers with symmetry mates (Sallai and Tucker, 2005). There are also a number of high-resolution crystal structures of the central domain of PspF, including the hexameric oligomer in complex with ATP analogues, produced by soaking the analogues into preformed crystals (Rappas, 2006).

Figure 1.5. Summary of current known structures of σ^{54} activators. Circular regulatory domains indicate two-component receiver domains. Star-shaped regulatory domains indicate GAF domains. Color coding is as follows: red indicates a high-resolution structure is available; green indicates multiple high resolution structures are available; blue indicates a low-resolution structure is available.



Recent studies further include the cryo-EM and SAXS reconstructions of the central and DNA-binding domains in the activated AlFx-bound state of NtrC (de Carlo, et al., 2006). However, no structure of a DNA-binding domain of a σ^{54}-activators in complex with DNA has been presented to-date. Structure determination of a DNA-binding domain complex from *A. aeolicus* has been my central thesis project, and is described in Chapter 2. Additional work presenting the structures of the DNA-binding domains of NtrC1 and NtrC2, and the comparison of these domains with others in the NtrC family, is the subject of Chapter 3.

1.7 Overview of two-component signal transduction.

The regulatory domain of NtrC-like transcriptional activators is a member of a two-component signal transduction system (TCS). TCSs are the primary prokaryotic regulatory networks that govern rapid physiological changes. TCSs have been observed in almost all bacteria, with an average of 52 TCSs per organism (Cock and Whitworth, 2007). TCSs can respond to various stimuli including temperature, pH, oxygen, pressure, osmolarity, auto-inducer compounds, the redox state of electron carriers, and contact with host cells (Beier and Gross, 2008). The number of TCSs encoded in a cell was found to roughly correlate with a cell's ability to adapt to diverse stimuli (Beier and Gross, 2008).

Since cell growth and pathogenicity are often regulated by TCSs, these pathways have received attention as antibiotic targets (Gotoh et al., 2010).

TCSs involve three steps: sensing of a stimulus, signal transmission, and a response, usually connected by a series of phosphoryl transfers. TCSs contain two component proteins, including a sensor kinase (SK), usually a histidine kinase (HK) in prokaryotes (a serine-threonine kinase in eukaryotes), and a response regulator (RR). SK's are multidomain proteins that usually consist of a sensor domain (most commonly a Per-Arnt-Single Minded domain) or a series of sensor domains, and the two-part "kinase domain" that consists of a Dimerization, Histidine, phosphorylation domain (DHp), and a catalytic ATPase domain (CA) that interact with an interdomain β-sheet. HKs have the ability to autophosphorylate the conserved histidine residue in the DHp region, and transfer it to the RR. In the majority of cases (~83%), HKs include a region of transmembrane helices and often transduce environmental signals across membranes. The remaining 17% of HKs are fully cytosolic. RR proteins, such as NtrC, usually consist of two or more domains, including a receiver domain, that receives the signal from the SK, and an output domain, that translates this signal into a functional change.

Due to the great diversity in sensor domains utilized, HKs are able to sense a wide variety of environmental signals. These signals modulate a SK's ability to autophosphorylate, and hence, to activate its associated RR. In the case of σ^{54} transcriptional activators, phosphorylation of the two-component receiver domain stimulates transcription.

1.8 Thesis outline

This thesis spans a wide range of topics in σ^{54}-dependent transcriptional activation. Chapter 2 describes my work determining the structure of the DNA-binding domain of NtrC4 from *Aquifex aeolicus* in complex with DNA by X-ray crystallography. Chapter 3 describes my work on the structures of the DNA-binding domains of NtrC1, NtrC2, and NtrC4, with commentary on the flexible linker between the central and DNA-binding domains and the nature of DNA-binding cooperativity between tandem binding sites. The work presented in both Chapter 2 and 3 are in preparation for submission to peer-reviewed journals. Chapter 4 describes my efforts to study the σ^{54} factor directly by NMR, and to elucidate the structural properties of Region $\sigma^{N}3$ in collaboration with the Darst lab. Chapter 5 describes my work towards a model for activator-gene associations of σ^{54} activators in *Aquifex aeolicus* using sequence analysis, gel-shifts assays, and a rigorous literature review. This chapter also describes the exploration of a new two-component system and includes the characterization of a heme-binding histidine kinase that activates NtrC3.

Chapter 2: DNA-recognition by a σ^{54} transcriptional activator from *Aquifex aeolicus*

2.1 Summary

The dimeric DNA-binding domain of NtrC-like activators in *Aquifex aeolicus* have not been studied structurally as much as the R and C domains. This domain exhibits a helix-turn-helix fold and high sequence similarity to the prolific DNA-bending protein, the Factor for Inversion Stimulation (Fis). In this chapter I describe my work to solve the structure of a protein-DNA complex of a DNA-binding domain from *A. aeolicus*. Complexes of the DNA-binding domains of both NtrC1 (1DBD) and NtrC4 (4DBD) with 21-mer DNA were analyzed by NMR spectroscopy and subjected rigorous optimization. However, although good HSQC spectra could be produced, all triple-resonance experiments gave poor data quality. Therefore, I switched techniques and solved the structure of the 4DBD in complex with its high-affinity DNA-binding site by X-ray crystallography. This structure, together with the structure of the free 4DBD solved by Johanna Heideker, demonstrates how 4DBD binds DNA and enables a comparison of 4DBD and Fis binding to DNA.

2.2 Introduction

2.2.1. Structures of σ^{54}-activator DNA-binding domains

A great deal of structural and biochemical work has been devoted to understanding the mechanisms for regulation and ATPase activity of σ^{54} transcriptional activators (Wigneshweraraj, 2008). However, less is known about activators' DNA binding domains, particularly bound to their target DNA. The structures of activator DBDs reported in detail to date are from NtrC from *Salmonella typhimurium* (Pelton,1999), although mutations used to stabilize the protein eliminated DNA binding capability; and from ZraR (Sallai, 2001). These have helix-turn-helix folds, buttressed by a third helix, as seen in many other bacterial DNA binding proteins (Aravind, 2005). In NtrC there is also a fourth helix that plays a role in dimerizing the DBD and holding the recognition helices with a spacing close to that of two successive major grooves of DNA. No structures of activator DBDs have been reported in complex with DNA.

2.2.2. NtrC and Fis are homologues

The DBDs of NtrC proteins are close homologues and evolutionary ancestors of the versatile DNA-binding and bending protein, Factor for Inversion Stimulation (Fis) (Morett & Bork, 1998). During exponential growth, Fis is one of the most abundant proteins in enteric bacteria, and is involved in a number of processes ranging from site-specific recombination, integration-excision reactions of phages, position and negative transcriptional regulation, and cell cycle timing in chromosome replication (Stella, 2010).

Fis also shares some similarity with DNA structural proteins like Integration Host Factor (IHF) in that it can nonspecifically bind DNA with high affinity. A number of structures have been solved for Fis alone (Cheng, 2000; Yuan, 1995; Kostrewa, 1993; Yuan 1993), and there is a multitude of biochemical data describing its binding to a variety of nonspecific DNA-binding sequences, and its potency in bending DNA up to 90º (Cheng, 2000; Pan, 1996; Thompson, 1988).

The similarity in sequence and structure of Fis and NtrC family DBDs opens the question of how these two protein families perform very different functions with different requirements for binding specificity. For NtrC, DNA-bending has been observed with electron microscopy (Revet, 1995), but no high resolution structural or biochemical studies have been performed to characterize the degree of bending. Additionally, there is no evidence that NtrC proteins bind DNA nonspecifically with high affinity or perform any of the additional chromatin-like DNA-restructuring functions of Fis. Therefore, the structural details that lead to the difference in behavior of these two close homologues will aid our general understanding of how proteins bind and bend DNA.

Since the time the original NtrC and Fis structures were solved, there has been an interest in understanding the details of their complexes with DNA. However, until recently, there had been one NMR study exploring backbone structure and residues involved in the DNA-binding of the σ^{54}-activator NifA (Ray et al., 2002), but no structure of a complex. The first high-resolution structures of Fis complexes were solved recently by Stella *et al.,* (2010). The structure of Fis bound to a high-affinity site, and bound to a number of binding site variants, provided characterization of protein-DNA interactions and the mode of DNA recognition. A model was presented in which Fis searches the DNA for regions with compressed minor grooves, then binds, exploiting the induced fit. The minimal reliance on specific base contacts and the large number of DNA backbone contacts explains the ability of Fis to bind without significant sequence specificity.

The structures of the free and bound NtrC4 DNA-binding domain (4DBD) allow a detailed comparison with Fis to understand the similarities and differences that modulate binding specificity and DNA distortion. These structures, and comparisons with other structures in the family, provide new insights into how σ^{54} transcriptional activators bind DNA.

2.3 Materials and Methods

2.3.1 Protein expression

Four DNA-binding domain constructs were cloned, including NtrC1-DBD (1DBD) (373-439 and 393-439) and 4DBD (365-442 and 375-442). All constructs were cloned into a Pet21a vector with ampicillin resistence (Michaeleen Doucleff; Ann Maris). Plasmids were transformed into TFB-chemically competent *E. coli* cells with a Rosetta.pLys plasmid. To express isotopically-labeled protein (either ^{15}N or $^{15}N/^{13}C$), the media-switching technique by Cai, *et al.,* (1998) was used. A single colony from a fresh transformation (frozen cell stocks were not effective) was added to 4 mL of LB supplemented with 100 mg/L ampicillin and 35 mg/L chloramphenicol, and was cultured overnight at 37ºC. One mL of starter culture was added to 4x1L of LB media, also supplemented with ampicillin and chloramphenicol. Cells were grown with shaking at

37ºC until an O.D. of 0.7 was reached. Cells were spun down for 20 minutes at 5000 rpm, washed with 1 L of M9 salts media (Cai, et al., 1998), spun down again for 20 minutes at 5000 rpm, then resuspended in M9 buffer containing 0.5 g NaCl, 6.78 g Na_2HPO_4, 3.0 g KH_2PO_4, 1 g $^{15}NH_4Cl$, and 2 g ^{12}C or ^{13}C-glucose. After resuspension in the final media, cells were shaken for 1 hour at 37ºC, then induced with 1 mM IPTG. Cultures were then grown overnight with shaking at 25ºC for approximately 16 hours, pelleted at 5000 rpm for 20 minutes, and resuspended in a resuspension buffer (50 mM sodium phosphate buffer pH 6.5, 400 mM NaCl, 1 mM EDTA).

2.3.2. Protein Purification

1DBD and 4DBD proteins were purified with identical procedures. Resuspended cell pellets were supplemented with PMSF and sonicated at 4ºC for 6x30 second intervals at high power. Lysate was heated at 80ºC for 30 minutes in a water bath to precipitate *E. coli* proteins, and ultracentrifuged for 30 minutes at 30,000 rpm. The supernatant was dialyzed overnight into a buffer with 50 mM Sodium phosphate pH 6.5, 200 mM NaCl, and 1 mM EDTA. Samples were then bound to a heparin column, and eluted with a linear salt gradient from 200 mM NaCl to 800 mM NaCl. 1DBD and 4DBD samples eluted between 500 and 600 mM NaCl, and were concentrated with Amicon centricon concentrators with 3,000 kDa MWCO. Typical yields were ~7 mg per 1 L M9 media.

2.3.3. NMR of 1DBD and 4DBD complexes

NMR experiments were tried on four different complexes: 1DBD-short (373-439) and 1DBD-long(393-439) + the Aq_1119 high affinity binding site (CTTTTGCCAAAATGGCAAAAG); 4DBD-short (375-442) and 4DBD-long(365-442) + the –145 binding site (TTGCAAATTTGCAAATGCATAA); and 4DBD-short + the –85 binding site (TTGGTAAAACTACAAAAGAGG). Experiments conducted include the ^{1}H-^{15}N-FHSQC (*fhsqc*; Mori, et al., 1995), HNCA (*hncagpwg3d;* Grzesiek and Bax, 1992), HNCACB (*hncacbgp2h3d*; Shan, et al., 1996), 3D-HSQC-NOESY (*noesyhsqcf3gpsi3d*; Palmer III, et al., 1991), 3D-HCCH-TOCSY (*hcchdigp3d*; Kay et al., 1993). Purity of DNA-binding sites was checked by 20% denaturing PAGE gels including 8 M urea. If the DNA strand was ~95% pure, it was used without further purification. Otherwise, DNA strands were purified by large gel electrophoresis, as described in section 2.3.3. DNA strands were annealed for 5 hours in a 2-L hot water bath in deionized water solution with 2 mM $MgCl_2$. Protein solutions, in 50 mM sodium phosphate buffer pH 6.5, 200 mM NaCl, and 1 mM EDTA were titrated into the DNA solutions to attempt to reduce irreversible aggregation, and samples were desalted to 20 mM NaCl in the same buffer.

Complex samples were optimized and checked by NMR in a range of 10ºC to 50ºC; 0-200 mM NaCl; in buffers HEPES, Tris-HCl, Sodium Phosphate, and PBS; in a pH range from 6.5-8.0; in 0 or 1 mM EDTA; 0-2% glycerol; with the additives 1 mM ZnCl, $CaCl_2$, or $MnCl_2$; protein concentrations ranging from 100 μM to 3 mM; and protein-DNA ratios ranging from 1:0 to 2:2.5.

2.3.4. Complex preparation for crystallography

DNA forward and reverse strands listed in Table 2.1 were ordered unpurified from the Keck Facility at Yale University for crystal trials. One brominated strand was ordered for MAD phasing, incorporating 5'-bromo-dU and 5'-bromo-dC bases into the strand. All DNA strands were purified separately by large gel electrophoresis in 17% 29:1 denaturing polyacrylamide gels in 8 M urea, at 700 V for ~16 hours. DNA bands were visualized by DNA shadowing, excised from the gel with a razor, pulverized by extrusion through a syringe tip, and soaked in 25 mL aliquots of TBE 1X buffer with rigorous shaking at 37C. A total of 150 mL of aliquots were taken over 36 hours. Extracted DNA was desalted and concentrated with Waters Sep-Pak desalting columns, eluted into 50% methanol, and dried in a speed-vac centrifuge. Strands were resuspended in 50 mM HEPES pH 7.0, 20 mM NaCl, 1 mM EDTA ("binding buffer"), quantified by UV absorption at 260 nm, combined in 1:1 molar ratios in the presence of 2 mM $MgCl_2$, and annealed in a hot water bath for 3 hours. The resulting DNA was combined in a 1:2 molar ratio with 4DBD in DNA-binding buffer, and run over a Superdex75 size exclusion column equilibrated with the same buffer to reduce excess amounts of free ssDNA, dsDNA, or protein. The eluted complex sample was concentrated with Millipore centricon concentrators to a concentration of 20-30 mg/ml, and stored at 4ºC. Complex samples were quantified by Bradford assay.

Table 2.1. DNA strands used for X-ray crystallography. Overhanging bases are orange. Mismatched overhanging bases are blue. In the Brominated strand, X=5'-bromo-dC; Z=5'-bromo-dU.

	FORWARD STRAND	REVERSE STRAND
22mer:	TTGCAAATTTGCAAATGCATAA	TTATGCATTTGCAAATTTGCAA
20mer:	TTGCAAATTTGCAAATGCAT	ATGCATTTGCAAATTTGCAA
18mer:	TTGCAAATTTGCAAATGC	GCATTTGCAAATTTGCAA
20-OH1:	TTTGCAAATTTGCAAATGCAT	AATGCATTTGCAAATTTGCAA
20-OH2:	CTTGCAAATTTGCAAATGCAT	GATGCATTTGCAAATTTGCAA
20-OH3:	ACTTGCAAATTTGCAAATGCAT	TGATGCATTTGCAAATTTGCAA
20-OH4:	CTTTGCAAATTTGCAAATGCA	AGTGCATTTGCAAATTTGCAA
Brominated strand:		AGZGXATZTGCAAAZTTGCAA

2.3.5. Crystallography of 4DBD-complex

X-ray data collection and refinement was done in collaboration with Dr. Artem Lyubimov. 4DBD was co-crystallized with dsDNA in 30 mM Na-cacodylate pH 6.0, 140 mM spermidine, 50 mM potassium thiocyanate, and 15% (poly)ethylene glycol mw 400, at 20°C, cryoprotected in 20% (w/v) xylitol and frozen in liquid nitrogen until use. X-ray diffraction data were collected at the Advanced Light Source (ALS) beamline 8.3.1 from a single crystal to 3.0 Å resolution at a wavelength of 0.92 Å over a 120º wedge with 1º oscillations. Data were integrated with DENZO and scaled with SCALEPACK as implemented in the HKL2000 suite of software (Olson, et al., 1997). Initial phases were solved by molecular replacement using the program PHASER (McCoy, et al., 2007). The starting model included two copies of the 4DBD monomer from the free-protein crystal structure, together with a model of ideal 13-mer B-form dsDNA (3D-

DART server, van Dijk, 2009). Remaining DNA nucleotides were modeled manually using Coot (Emsley and Cowtan, 2004), and refined with PHENIX (Adams, et al., 2010). The structure was validated using SFCHECK (Vaguine, et al., 1999), PROCHECK (Laskowski, et al., 1993) and MOLPROBITY (Chen, *et al.,* and Richardson, 2010). The final data processing and structure refinement statistics are summarized in Table 2.3.

2.3.6. Binding affinity assays

Binding affinity assays were used to determine the equilibrium constant of dissociation (Kd) of 4DBD to a number of binding sites. DNA was ordered from Integrated DNA Technologies with a 5'-fluorscein tag on the forward strand listed in Table 2.1. Forward and reverse strands were annealed for 5 hours in the presence of 1 mM MgCl2 and stored at 4ºC away from light. Protein-DNA aliquots were mixed with 10-15 nM DNA in each track, increasing protein concentrations in the range of 0-400 μM, 25 mM NaCl, 50 mM Tris-HCl pH 8.0, and 15% glycerol. Samples were immediately run on 20% native acrylamide gels for 2 hours at 80 V, then imaged with a Typhoon. Band intensities were quantified by QuantityOne Software, and plotted with Origins 8.1. Curves were fitted with non-linear curve fits based on a two-state saturation binding model.

2.4 Results

2.4.1. NMR of 1DBD and 4DBD complexes

The HSQC spectra of the 1DBD complexes show considerable line-broadening in the central region of the spectrum, indicating a chemical exchange process (such as protein association and dissociation from DNA), a dynamic process (such as a wobbling motion of each monomer on the DNA), or aggregation. The best HSQC spectra could be produced in the 4DBD-short complex, so subsequent studies focused exclusively on 4DBD complexes. However, although the 4DBD-complex produced moderately high-quality HSQC data and moderate-quality HNCA data, other 2D and 3D experiments (TOCSY, NOESY, HNCACB, HBHACONH) gave reproducibly poor data quality. This is highly unusual: it is a common observation in NMR that samples that produce reasonably high-quality HNCA spectra generally are able to produce legible TOCSY and NOESY spectra. It was found that this problem was also exhibited by the free 4DBD-short and 4DBD-long constructs, in the absence of DNA. Since 4DBD is a relatively small dimeric protein (17 kDa) that is very stable, very soluble, and produces high-quality HSQC spectra, this observation was surprising. A summary of NMR experiments performed is listed in Table 2.2.

Subsequently, the 4DBD-long and -short constructs and the 4DBD-complexes were exhaustively optimized based on protein concentration, protein:DNA ratio, pH, temperature, buffer, salt, purification method (gel-purification of DNA, size-exclusion purification of preformed complexes), and additives (EDTA, Zn^{2+}, Ca^{2+}, Mn^{2+}). The best conditions were found to be 50 mM HEPES buffer pH 7.0 and 10 mM NaCl (complex) or

200 mM NaCl (free protein), and 1 mM EDTA, with spectra taken at 313-323 K. Free protein samples also exhibited a notable improvement in the NOESY spectra after refolding in 6 M guanidine solution, and rapid dilution into the above sample buffer, in the absence of EDTA. In case the asymmetry of the binding site was causing the spectral broadening, I also considered monitoring the spectrum of 4DBD in complex with a shortened, palindromized-form of the binding site (TTTGCAAATTTGCAAA). However, gel-shift assays indicate that 4DBD does not bind to this site, most likely due to its length.

One possible explanation for poor data quality is the presence of an aggregation process. If 4DBD is able to form nonspecific higher-order oligomers on the DNA, this may cause the tumbling of the sample to be just slow enough to cause line-broadening in the triple-resonance experiments. The presence of aggregation is also suggested by a number of other techniques: in gel-shift assays as well as fluorescence-anisotropy experiments, samples with high protein concentrations exhibited additional bands and a gradual rise in correlation time, respectively, which suggests an aggregation process occurs above 0.4 μM. Also, $^{1}H/^{15}N$-HSQC-spin-echo NMR experiments, which can be used estimate the molecular weight of a sample based on their T2 values, found that the sample tumbles at rate corresponding to a molecular weight near16 kDa at 323 K, and 34 kDa at 298 K. This suggests the presence of tetramers at lower temperatures. Indeed, free 4DBD was soluble in 0.5 mM solutions at low salt concentrations (~10 mM) above 40 ºC, but crashed out of solution immediately if the temperature were lowered. On the other hand, if the sample were forming tetramers, it should be possible to produce high-quality NMR data using partial or uniform deuteration of the sample, combined with TROSY spectroscopy, as described in detail in Chapter 4. To test this hypothesis, a 70%-$^{2}H/^{15}N/^{13}C$ sample of 4DBD was prepared and analyzed by NMR, and found to provide no improvement in the data quality of NOESY, HNCACB, TOCSY spectra. A second possible explanation of the poor data quality is the presence of a chemical exchange process. For example, if oligomers were able to form and dissociate on an intermediate exchange timescale, this would cause another form of line-broadening, which could stem from a relatively small population of oligomers in solution.

Both exchange and size considerations may be contributing to line-broadening. However, since the causes of line-broadening are not easily quantifiable by available techniques, and may not be avoidable in solution, I decided to discontinue NMR analysis of these systems.

2.4.2. Crystal structure of free-4DBD

The unbound crystal structure of 4DBD-long was solved previously by Johanna Heideker and published together with structures of other NtrC4 domains (Batchelor, et al., 2008). However, the structure was included in that paper to provide evidence for DNA-binding domain dimerization, and was not discussed in any detail.

The 4DBD crystallized with two dimers in the asymmetric unit. In keeping with helix-turn-helix nomenclature, each monomer consists of four helices, A-D, with the D helix representing the "recognition helix" that is typically inserted into the major groove of DNA (see Figure 2.1A). The A and B helices of 4DBD form an antiparallel four-helix bundle with the corresponding helices in the adjacent monomer. As already observed for

the DNA-binding domain of NtrC (Pelton et al., 1999), the A helices do not overlap significantly but pack against the long B helix (Lys29 – Tyr47) through hydrophobic sidechain interactions. Helices C (Leu51 – Ile58) and D (Leu62 – Leu72) form a classic helix-turn-helix DNA binding motif at the C- terminus of the molecule. The HTH motif is extended by a tail consisting of residues Leu72 to Ser79. Dimer formation buries an area of 2160 Å2, corresponding to 30% of the total surface area of 4DBD.

Table 2.2. NMR experiments performed and relative data quality obtained.

Sample	DNA	NMR experiments	Data quality	Comments
1DBD-short	-	HSQC, HNCA, HNCACB, 3D-HSQC-TOCSY, 3D-HSQC-NOESY	Good	Performed by Dr. Ann Maris.
1DBD-short	+	HSQC, 3D-HNCA	Poor	Severe line-broadening
1DBD-long	+	HSQC, 3D-HNCA	Poor	Severe line-broadening
4DBD-long	-	HSQC, 3D-HNCA, 3D-CBCACONH	Good	Performed originally by J. Heideker; repeated
	-	3D-HSQC-TOCSY, 3D-^{15}N-TOCSY, 3D-HCCH-TOCSY, 3D-HSQC-NOESY	Poor	Severe line-broadening
	+	HSQC	Good	
	+	HNCA, CBCACONH	Moderate	some line-broadening, low intensity, missing peaks
4DBD-short	-	HSQC, HNCA, CBCACONH	Good	
	-	3D-HSQC-TOCSY, 3D-HSQC-NOESY, 3D-HBHACONH, ^{15}N-TOCSY	Poor	Severe line-broadening
	+	HSQC, HNCA	Moderate	Some line-broadening, missing peaks
		CBCACONH, HNCACB, 3D-HSQC-TOCSY, 3D-HSQC-NOESY, 3D- HBHACONH, ^{15}N-TOCSY	Poor	Severe line-broadening

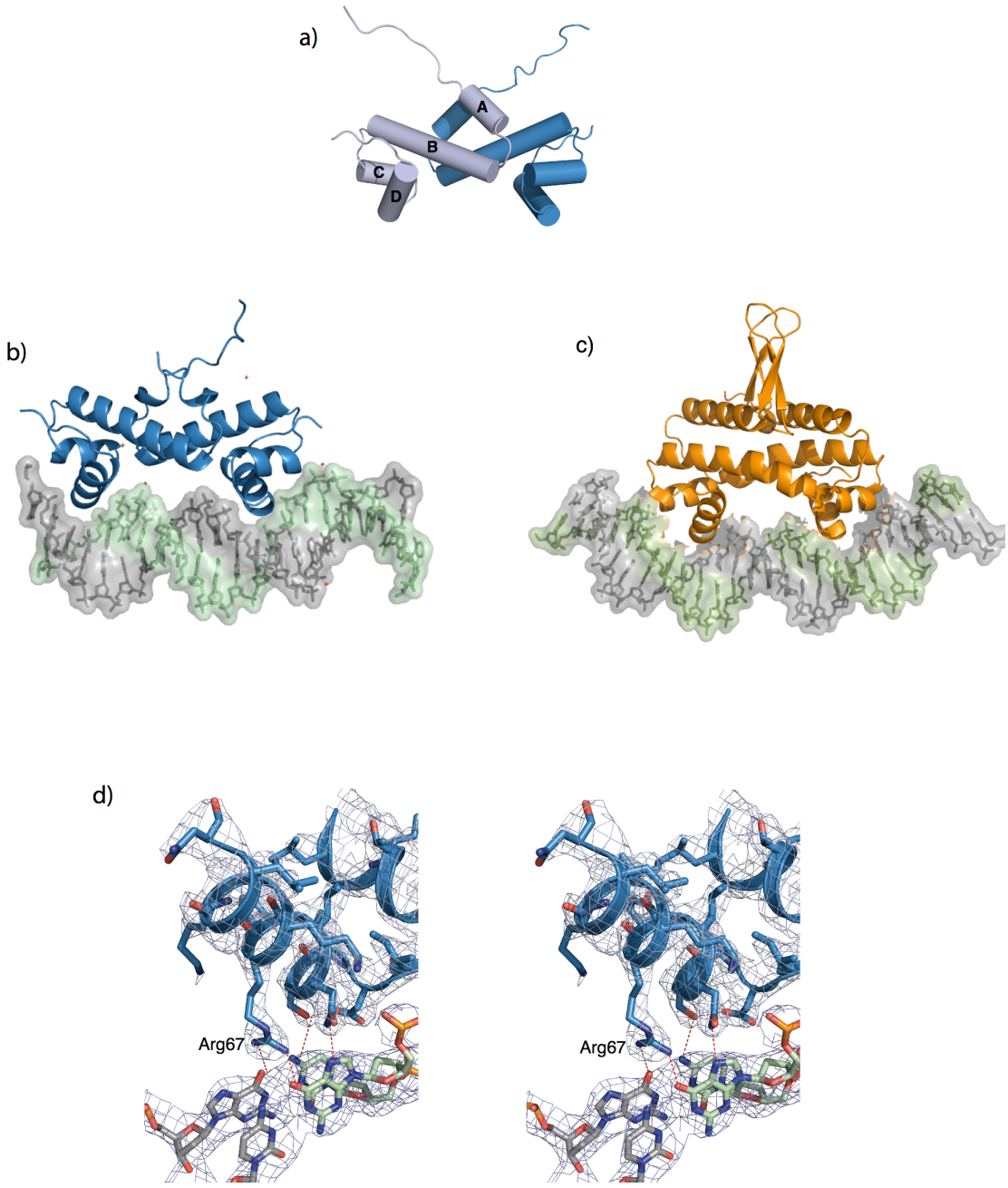
Table 2.3. X-ray crystal structure statistics of the 4DBD *free structure and complex.*

	4DBD	4DBD complex
Data Collection		
Space group:	$P2_1$	$P2_1$
Unit Cell parameters		
a, b, c	42.2, 55.7, 62.3	67.4, 28.8, 74.9
α, β, γ		90, 94.6, 90
Resolution (Å)	41.17-2.25 (2.33-2.25)	48.1-3.0 (3.1-3.0)
Unique Reflections	13,545	5,902 (454)
Completeness (%)	98.2%	96.8% (76.2)
Multiplicity		3.3 (2.3)
I/σ_I	16.5(3.9)	15.5 (2.5)
R_{merge}		0.063(0.31)
Refinement		
Resolution limits (Å)		48.1-3.0 (3.1-3.0)
R_{work}/R_{free} (%)	20.3 / 24.3	24.3 / 26.4
Number of atoms		
Protein		1106
DNA		861
B-factors		
Protein		95.4
DNA		112.6
RMSDs		
bond lengths (Å)		0.004
bond angles (°)		0.802

No density could be observed for residues Met1 – Trp16 and only weak density was visible for Leu77-Ser79, indicating that the two termini are disordered or flexible.

The overall fold is similar to the DNA binding domains of Fis (Fig. 1b) and NtrC. Superposing the three C-terminal helices (residues Lys29 – Leu72) with the corresponding residues in Fis and NtrC using the program Superpose (Krissinel et al., 2004) gives root mean square deviations of 0.65 Å and 2.1 Å, respectively. Dimerization occurs mainly through hydrophobic interactions between the A and B helices. Contacts between the end of the B helix and the loop between helix C and D may also contribute to dimer stabilization. Close intra-monomer contacts are observed between the sidechains of Leu22 and Leu19 in helix A and Ile38 of helix B. Close inter-monomer contacts are made between the sidechains of Leu19, Leu23 and Ile38. The B helices of each monomer cross at a 40º angle such that the phenyl rings of residue Phe35 of each monomer stack upon each other in a parallel displaced manner. The two Phe35 residues form a point of two-fold symmetry.

Figure 2.1: Structure of free and bound forms of 4DBD. *a) Cylinder view of free* 4DBD *crystal structure showing the two symmetrical monomers in light and dark blue; helix labeling convention A-D is shown. b) Cartoon view of* 4DBD *bound to site_1 DNA. Forward strand is gray, reverse strand is green, c) structure of Fis bound to DNA (Stella, 2010), d) stereo image cutaway of structure and crystal density highlighting specific contacts.*



2.4.3. Determination of the DNA-binding site for NtrC4

Gene-association studies, which find that NtrC4 binds to the UAS of the *lpxC* gene in *A. aeolicus*, are described in detail in Chapter 5. To determine the specific binding site, Dnase I footprinting was performed on the *lpxC*-UAS by Dr. Yixin Huo. Two areas of significant protection were visible centered around -145 and -85 bp upstream of the transcription start site at a protein concentration of 10 nM. To further confirm these binding sites, the two sites and a series of mutant sites were purchased with a 5'-fluorescein tag and analyzed by fluorescence polarization and affinity binding assays. Since fluorescence polarization studies showed a steady rise in molecular weight with increasing protein concentration, most likely due to aggregation, K_d values were calculated based on affinity binding assays. Resulting K_d values are summarized in Figure 2.2.

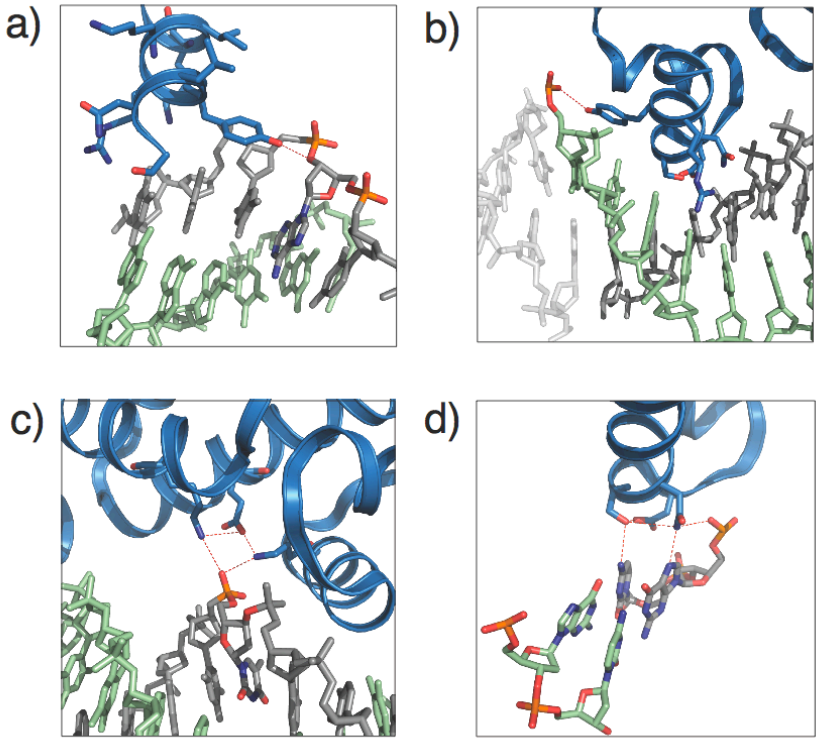
2.4.4. Structure determination of the 4DBD-site_1 complex

Since flexible regions can be detrimental to crystallization, the 4DBD-short construct, which removes 9 flexible residues at the N-terminus, was used for complex formation. The 4DBD complex crystallized in $P2_1$ symmetry and was solved to 3.0 Å resolution with 22mer DNA that contained a two-base-pair mismatched overhang (Forward 5'-**AC**TTGCAAATTTGCAAATGCAT-3'; Reverse 5'-**TG**ATGCATTTGCAAATTTGCAA-3'). Due to the DNA-mismatches in the overhangs, the first (5')-adenine base of the forward strand and (5')-thymine base of the reverse strand flipped out of the helix, and the resulting one-base-pair G/C overhang self-annealed, forming a continuous double helix through the crystal lattice with 21 base pairs. Although a variety of sequences of different lengths and varying overhangs crystallized, these samples gave poorer resolution (>4Å). Since the flipped-out bases appear to be disordered in the structure and are not clearly visible in the density, the numbering convention used in this paper will describe only the 21 nucleotides observed (see Figure 2.1).

2.4.5. 4DBD-complex: protein-DNA interactions

As in the free protein, the N-terminal residues preceding Arg16 (residues 10-15 in the short construct) form what appears to be a flexible linker to the NtrC4 central ATPase domain. The amino acid sequences for this segment of the protein in various NtrC family members show low sequence conservation, and are variable in length. Omitting this region, the free and bound forms of 4DBD are very similar in structure, with an overall RMSD of 0.57 Å. The most notable differences relative to the free 4DBD structure are the orientations of the sidechains of R67 and Y66. Compared to the free protein, Arg67 extends outward towards the DNA in the complex structure, and Y66 rotates by ~ 90º to contact the DNA backbone.

Figure 2.2. Cutaway views of nonspecific interactions: a) Y66 interaction with internal backbone residue, contrasted with b) the other monomer's Y66 interaction with the first resolvable base pair backbone near the nicked DNA (neighboring DNA symmetry mate shown in white), c) hydrogen bonding ring at ±3Thy, d) Asp61, Ser63, Asn64 hydrogen bonding network.



The structure of the bound protein, and expansions showing some of the protein-DNA interactions are presented in Figures 2.1 and 2.2. As is expected for helix-turn-helix proteins, the majority of protein-DNA interactions occur in the D helices (Aravind, *et al*., 2005). In this thesis, the bases in the binding site are labeled from –8 to +12, centered around the palindromic center of the 14-base pair "core" binding site (see Figure 2.3). Since the protein is a symmetric dimer, most interactions with one half-site binding sequence are repeated in the other half-site. For simplicity, interactions in only one end of the helix will be described below, except in the cases where asymmetry exists in the protein-DNA interactions.

The primary sequence-specific contacts are mediated by Arg67 and Ser63. The two sidechain NH_2 groups of Arg67 form hydrogen bonds with the O6 atoms of +4G(F) and +5G(R), bridging the two strands of DNA. The hydroxyl oxygen of Ser63 accepts a hydrogen bond from the amino group of +5C(F), which is buttressed by a hydrogen bond to the sidechain carboxyl group of Asp61 (See Fig 2d). Asn64 also forms a hydrogen bond with the N7 group of +4Gua(F), forming a contact that selects for a purine (Fig 2d). In addition to these hydrogen bonds, Asn64 is held in place by van der Waals contacts to

Arg67 and +4Gua phosphate backbone. This combination of a hydrogen bond network and van der Waals contacts produces a structurally well-defined protein-DNA binding interface at the D helices.

There are also a number of nonspecific contacts with the DNA phosphate backbone. To orient the recognition helices in the major groove, Lys68 hydrogen bonds with the phosphate group on the +3T(F) backbone. Lys70 hydrogen bonds with the phosphate group of +7T(R). A hydrogen bond is also formed between the backbone amide of Asp61 and the phosphate of G12(F).

The only other nonspecific contacts outside of this region are a hydrogen-bonding ring between residues in both A and B protein monomers that form the following array of hydrogen bonds with a thymine phosphate: 17Thy(R) -- 32Lys(A) -- 36Glu(B) – 68Lys(B) – 17Thy(R) (Figure 2.2c). All of these interactions are repeated in the second subunit. The only notable difference between the two subunits is the conformation of the Tyr66 residue, which forms a hydrogen bond with the phosphodiester O on one side of the monomer [+8T(R)], but bonds directly with the final free phosphate on the other monomer [1C(F)], due to the asymmetrical presence of the DNA nick on the 5'F end (Fig 2a, b).

2.4.6. 4DBD-complex: DNA conformation overview

The structure of the DNA in the complex shows only slight deviations from B-form DNA. Although the two protein subunits overlay closely, the fact that the actual binding site is not at the center of the oligonucleotide that crystallized causes slight structural differences at the two ends. On the 5'-forward end, the DNA nick occurs only two base pairs away from the -5G/-4C base pairs (the "GC motif"). On the 3'-forward end, there are an additional 9 base pairs following the GC motif. This asymmetry is reflected in many features of DNA distortion presented below.

The DNA-binding site has an A-tract separating the GC motifs that are sequence-specifically contacted, a sequence motif that is relatively common as a target for helix-turn-helix DNA-binding proteins. It has been shown that regions of four or more adenine residues in a row form somewhat unusual structures that are relatively inflexible. A variety of different studies have shown definitively that periodic A-tracts can form stably curved DNA, although the precise basis for these features are still debated (Haran, 2009). Pyrimidine(Y)-Purine(R) base steps have been shown to be more deformable than YY, RR, or RY steps, and A-tract sequences coupled with YR steps are one of the most common elements in sequences that undergo distortion by proteins (Beveridge, 2004; Olson, 1998). The "TGCA" motif that repeats in the binding site for 4DBD provides two points of flexibility (T-G and C-A), enabling the compression of the minor groove and expansion of the major groove. Plots of DNA roll and bend per base, and major/minor groove widths are presented in Figure 2.5. The DNA in the 4DBD complex exhibits an overall bend of ~9°, with the majority of bending occurring at the 3'F-terminal A-tract base pairs +5 to +7.

Most of the DNA is close to B-form duplex; however, there are two base pairs at +8A(F)-+8T(R) and +12T(F)-+12A(R) that show significant shear. The first is in the A-tract following the second GC pair, which is also the region of maximal DNA bending.

The second instance is in the final base before the nick, and may be the product of distortion due to crystal contacts since it is far from protein contacts.

2.4.7. Biochemical analysis of DNA-binding site of NtrC4

To probe the relative importance of different base pairs in the DNA-binding site, gel-shift assays were performed on fluorscein-labeled DNA strands with changes introduced at specific base pairs (Figure 2.4). A full 26-bp site that includes the core palindromic binding sequence flanked by 8 additional base pairs on either end was determined to have a K_d of 1.31 +/- 0.36 nM. As expected, a single mutation of the specific contact base pairs G or C to T reduced binding affinity 80-to-120-fold (site_2, site_3, site_4, and site_5). A disruption of the A tracts that are peripheral to the GC sequences caused a ~100-fold reduction in binding affinity (site_6), suggesting the importance of the peripheral A-tracts and DNA flexibility in these regions. In contrast, the insertion of CG bases in the middle of the central A-tract (site_8) has a more subtle effect on binding, reducing affinity ~30-fold. The binding curves for site_3 and site_4 exhibit no binding at low concentrations (<100 nM), followed by sudden binding at higher concentrations (~100 nM), suggesting that they are not governed by a two-state model. Therefore, the binding to site_3 and site_4 are not treated quantitatively, but suggest a lower binding affinity than the first site occupied, and possibly cooperative binding.

2.5. Discussion

2.5.1. The relevance of lpxC activation to NtrC4 activity

The lpxC gene is involved in the biosynthesis of Lipid A, the hydrophobic anchor of lipopolysaccharide, which forms the outer monolayer of the outer membrane of most gram-negative bacteria (Raetz and Whitfield, 2002). Since this outer membrane in gram-negative bacteria is implicated in numerous forms of antibiotic resistance, it has sparked interest as an antibiotic target (Barb, 2008). The finding that NtrC4 regulates lpxC expression appears to be the first identification of σ^{54} control of cell-surface lipid metabolism genes. In our previous work on the regulation of NtrC4 we found that it can assemble to an active oligomer even in the absence of phosphorylation (unlike homologs NtrC1 in *A. aeolicus* and DctD in *S. meliloti*). This may be consistent with a low level of constitutive transcription of the lpxC gene, with cell cycle dependent transcription induced by phosphorylation.

2.5.2. Comparison of 4DBD free protein to Fis

It is informative to compare the structures of the free Fis and 4DBD. One important difference between these structures is the spacing between the recognition helices, which is 24 Å in 4DBD, and 20 Å in Fis. Without protein or DNA distortion, the 4DBD D helices are able to fit into sequential major grooves of ideal B-form DNA (gap ~

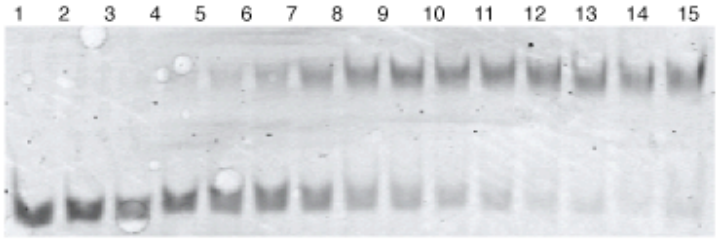
21 Å), whereas Fis would not be able to fit into neighboring major grooves without compression of the straddled minor groove (Hengen, 1997). It is not surprising, therefore, that 4DBD produces a much subtler bend in the DNA than Fis. Interestingly, the DNA-binding domain of another homolog, NtrC from *Salmonella enterica serovar typhimurium*, has an inter-D-helix spacing of only ~14 Å (Pelton, 1999). Although no structure of a complex with DNA is available for it, there is evidence for significant induced DNA bending by NtrC.

2.5.3. Comparison of the 4DBD and Fis complex structures: protein interactions

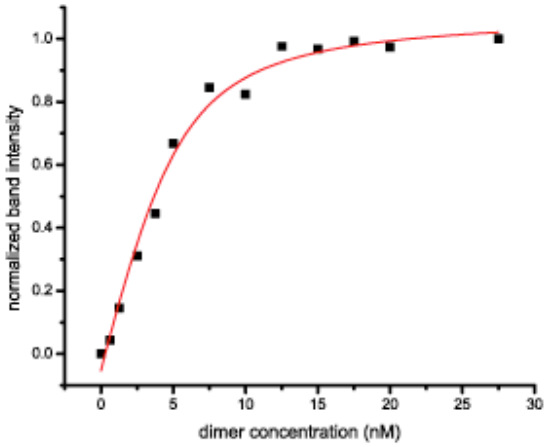
The DNA-binding preferences of Fis have been studied extensively (Keeler, 2005; Robinson, 1998; Ussery, 2001; Hengen, 1997, Grainger, 2006). Although Fis is known to bind to DNA sequences nonspecifically with nanomolar affinity, certain sequences can form stable sequence-specific complexes, characterized by tighter binding constants, a reduced binding dependence on the length of the DNA strand (if longer than the binding site), and the ability to form clear Dnase I footprints (Betermier, et al., 1994). This reflects Fis's variable functions, ranging from chromatin-like DNA-remodeling (Ussery, 2001; Blot *et al*.2006), to site-specific transcriptional repression (Ninnemann, et al, 1992). This spectrum of nonspecific and specific binding modes complicates the definition of a consensus sequence for Fis. This difficulty is illustrated by the conflicting predictions of 6,000-68,000 Fis-binding sites in the *E.coli* genome using different prediction methods (Ussery 2001, Hengen 1997). A recent genome-wide analysis of Fis-binding in Escherichia coli found 894 experimentally-determined Fis binding sites *in vivo*, and presents a revised assessment of Fis binding preferences (Cho, 2010). That work identified two high-affinity binding site consenses: one that includes a core nonpalindromic A-tract, and one that includes a core palindromic AT-tract. Both versions of the site have highly conserved G and C bases at –7 and +8 positions, respectively. It is interesting that although Fis, like NtrC4, is a symmetrical dimer, Cho and colleagues find that the nonpalindromic site has marginally greater binding affinity than the palindromic site. Since the A6 track is also known to produce the greatest intrinsic curvature to DNA segments (Koo, 1986), this concurs with the model presented by Stella, et al., that Fis selects for DNA-binding sites based on a compressed minor groove, binding with a further induced fit. Other residues in the Fis binding site are not highly conserved, but there is a notable bias against a thymine residue in the +4 position in both palindromic and nonpalindromic versions of the site (Cho, 2008).

Figure 2.4: a) Example titration gelshift determining K_d of 4DBD *in complex with site_2 (high affinity site). Each track contains 5 nM fluorscein-labeled DNA and the following dimeric protein concentrations: 0, .6, 1.3, 2.5, 3.8, 5.0, 7.5, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5 nM. b) plot of intensity vs. protein concentration from gel in a). Red line is nonlinear curve fit of data to the standard Adair two-site binding curve. c) Schematic of promoter and UAS and summary of binding site mutations tested and their resulting K_d's. *'d samples indicate non-two state behavior, as described in 2.4.6.*

a)



b)



c)

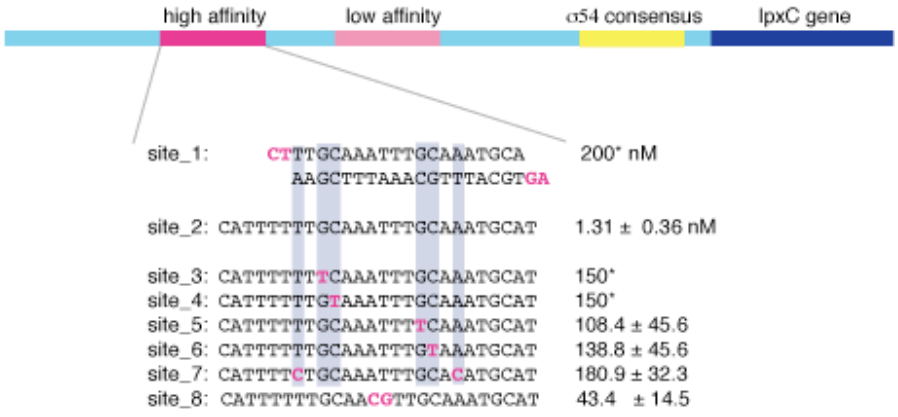
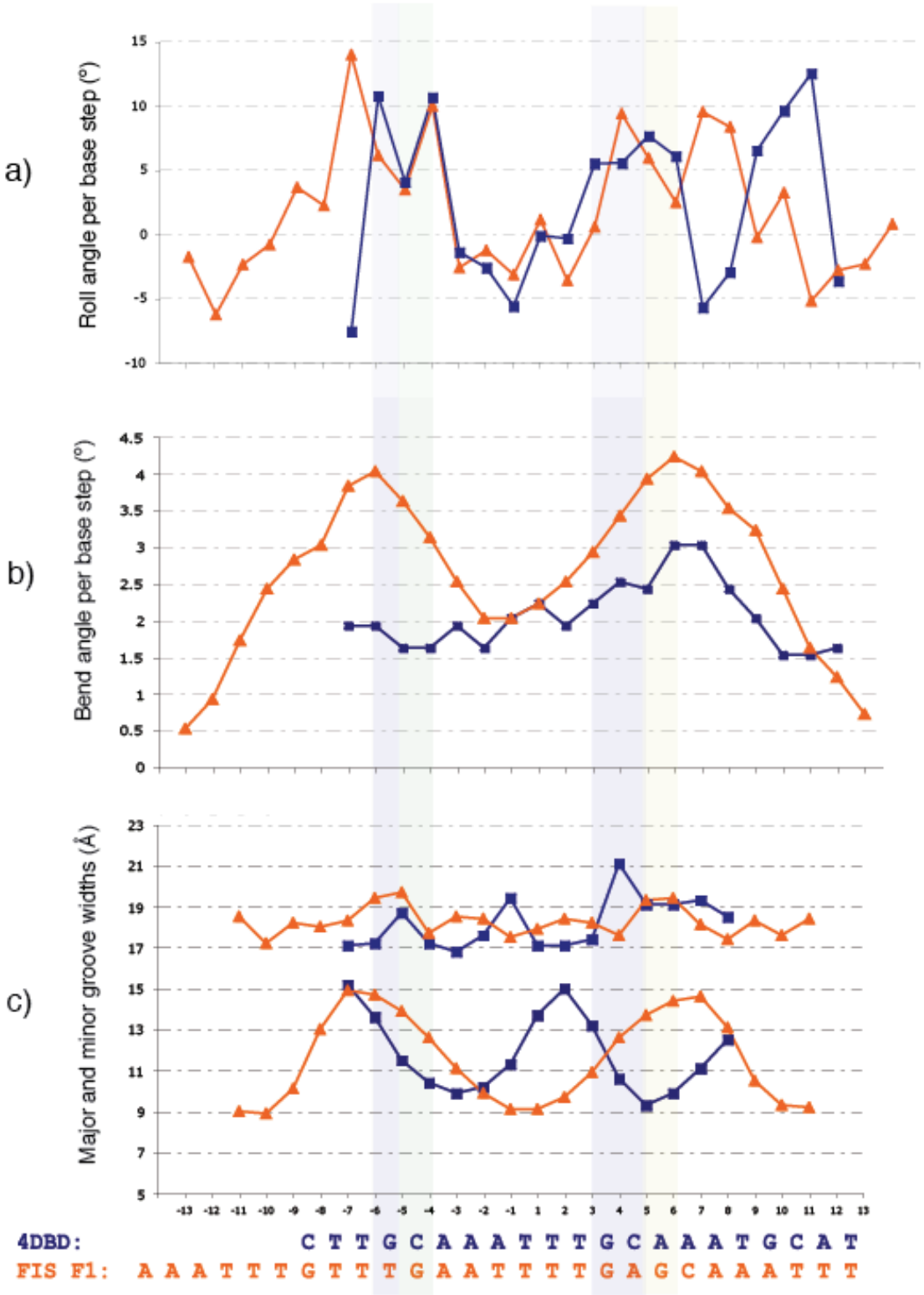


Figure 2.5. Comparison of DNA distortion in Fis (orange) and 4DBD*(blue). a) Roll angles, b) bend angle, c) major groove (upper lines) and minor groove (lower lines) widths per base step.*



The interactions between protein and DNA in 4DBD and Fis have a number of key similarities and differences. Both systems use an arginine residue (Arg67 in 4DBD, and Arg85 in Fis) as the dominant specific contact to a guanine base. Both also have an asparagine residue near the N7 group of a pyrimidine. In the case of Fis, Asn84 forms a hydrogen bond with +4G. Mutation of +4G to a C removes this hydrogen bond, but does not have a dramatic effect on binding affinity. However, mutating +4G to a T causes a steric clash that reduces overall binding affinity ~2500-fold (Stella, 2010). Therefore, it appears that the Fis Asn residue serves to prevent binding at sites with a ±4Thy residue. The same may be true for 4DBD, which has an asparagine residue (Asn64) that appears to be close enough to form three hydrogen bonds: one to the N7 pyrimidine group of +4G (2.9Å N-N distance), one to the phosphate backbone of +4G (3.0 Å N-O), and one to the hydroxyl group of Ser63 (3.8Å N—O). Higher resolution data would be necessary to precisely outline the interactions of this residue and determine whether it forms weak hydrogen bonds or strictly van der Waals interactions. However, it is apparent that if the ±4 residue were a thymine, the 5' methyl group would produce a steric clash with Asn64 that would either require a rearrangement of the residue or would reduce binding affinity. This may be a second cause, together with the loss of a specific contact to the +4G base, for the reduction of the 4DBD binding affinity ~15 fold in a +4G->+4T mutation (see figure 2.4, site_4).

2.5.4. Comparison of the 4DBD and Fis complex structures: DNA distortion

In Fis, direct interaction of arginine residues with guanines causes kinks with high roll angles (13º, 9º) at the GA steps. Although 4DBD bends the DNA much less than Fis, the roll angle profile for 4DBD is quite similar (see Figure 2.5). In 4DBD, high roll values occur at the YR steps surrounding the G basepairs ([–6T]-[–5G] and [–4C]-[–3A]). The bend angles per base pair were calculated with Curves+ (Sklenar, et al., 2009) and are compared in Figure 2.5. The 3'F end of the DNA shows a gradual curve at the peripheral A-tract by approximately 2º per base pair, in a pattern very similar to Fis.

In spite of these similarities to Fis, however, there is a striking difference in the major/minor groove widths. Stella, *et al.,* describe that the DNA in the Fis-DNA complex has expanded minor grooves opposite the major grooves where the interaction takes place, and a compressed central minor groove. They also find that the major groove widths are more or less constant, allowing the minor groove compression to lead to a global bend in the DNA. 4DBD also has only minor variations in major groove widths. However, the minor-groove compression profile for 4DBD appears to be inverted relative to Fis; the central minor groove is actually slightly expanded relative to B-form DNA, and the minor grooves opposing the major groove interaction regions are slightly compressed (Figure 2.5c).

This may be a reflection of the differences in spacing between specific contacts in the Fis and 4DBD binding sites and the relative positioning of specific and nonspecific contacts (Figure 2.3). Whereas 4DBD has 6 base pairs between the two specifically-contacted GC base pairs, Fis has 13 base pairs between the specifically contacted guanines. In 4DBD, the Arg67 residues point inward, toward the center of the binding site, to contact the GC motifs. In Fis, the Arg85 residues point outward, almost a 90º

rotation from the Arg67 position in 4DBD. The other specific contact in 4DBD, Ser63, has no equivalent residue in the Fis complex. However, there is close structural overlap between the following nonspecific contacts: Asn64, Asp61, and Tyr66 from 4DBD are in similar positions as Thr87, Asn84, and Arg89, respectively. Overall, the specific binding of Fis is predominantly "reaching outward", pulling the edges of the major groove inward, causing the compression of the central minor groove, and expansion of the peripheral minor grooves, with associated bending. 4DBD, by contrast, has specific contacts that "reach inward", stabilizing an expanded central minor groove and compressed peripheral minor grooves, not requiring significant bending.

In 4DBD, the minor groove expansion is greatest in the TTT base pairs at [+1]-[+3]. The greatest minor groove compression occurs at the CA steps in both [–4]-[–3] and [+5]-[+6]. The [+6]-[+7] base pairs directly following the [+5]-[+6] CA step reflect some of this distortion with elevated twist values, averaging 39º. The other CA step, at [–4]-[–3], also experiences an elevated twist value (39º).

The asymmetry of the DNA binding site and presence of DNA crystal contacts suggest that not all of the DNA distortion described above can be attributed to protein-DNA interactions. However, some key observations may be made: it appears that 4DBD compresses the central minor groove while expanding it in the peripheral A-tracts. The central A-tracts are fairly straight, but the CA base steps at [-4]- [-3] and [+5]-[+6] are sites of the most deformability and are overtwisted with high roll angles, generating a gradual bend in the 3'F peripheral A-tract. This combination of DNA base pair conformations produces a 9º overall global bend. Given the extended length of DNA at the 3' end, we would predict that the DNA distortions occurring at the 3'F end are the most illustrative of native binding. And owing to the symmetry of the protein and palindromic binding site, we would predict that if the complex were crystallized with fewer inter-DNA crystal contacts and with a longer DNA-binding site at the 5'F end, the bend seen at the 3'F end of the complex would also occur in a similar form on the 5'F end, producing a slightly more pronounced global bend, but still much less than that induced by Fis.

2.5.5. Conclusions on the mode of NtrC4 DNA-binding/bending

Stella *et al.,* describe that Fis binds DNA by scanning sequences for narrow minor grooves, binding sequences that have a loosely defined consensus sequence, and then inducing a bend in the DNA to generate a tight fit. This explains the nature of Fis as a DNA-bending protein that can bind a wide variety of nonspecific sequences with high affinity and induce different levels of bends. Although NtrC DNA-binding domains have been shown to bind some "governor sequences" nonspecifically to prevent DNA looping and inhibit transcriptional activation, they do so only at high concentrations and with very low affinity (K_d > 650 nM) (Ninfa, Reitzer, Magasanik 1987). There is currently no evidence that NtrC-like proteins bind DNA nonspecifically at high affinity.

This study provides an explanation for these differences. In addition to the differences in DNA distortion described above, Fis contains dramatically more nonspecific contacts to the backbone (18) than NtrC(10), and only one specific contact per half-site as opposed to three in NtrC4 (including the bidentate Arg67 interaction). One of the most important residues for nonspecific interactions in Fis, Lys90 (Skoko,

2006), has no structural equivalent in 4DBD. Spatially, nonspecific backbone contacts in Fis contact both sides of the major groove around the specifically bound G base pairs, whereas in 4DBD all nonspecific contacts occur on one side of the major groove, with the exception of Y66. This weaker interaction with both strands of DNA is likely to reduce 4DBD's ability to bend the DNA. It is possible that the lack of DNA-bending seen in NtrC4 is a reflection of the thermophilic nature of *A. aeolicus*. A number of indicators suggest that DNA-binding in thermophiles would reflect a different thermodynamic landscape than in mesophiles. For example, it has been shown that at higher temperatures curved DNA segments with periodic A-tracts straighten (Diekmann, 1987). It has also been suggested that a different balance of topoisomerases and DNA-binding proteins is required to keep the DNA annealed above its melting temperature (Guagliardi, 1997). Finally, it has also been observed that thermophiles tend to exhibit positive supercoiling of DNA, as opposed to the more common negative supercoiling observed in eukarya and eubacteria, which may effect the thermodynamics of DNA binding (Napoli, 2002). There is currently little understanding of how these features might affect the DNA-binding behavior at high temperatures in *A. aeolicus* cells.

NtrC proteins have a very different function in the cell compared to Fis, and appear to reflect these differences in their DNA binding. Sigma54 activation uses an ATP-energy requiring mechanism to tightly control transcriptional activation, and it is highly conserved in a number of very disparate families. Such stringent mechanistic conservation combined with the function of tight regulation suggests that a "loosening" of regulatory control, such as the potential of NtrC proteins to bind sequences nonspecifically, would be counterproductive. It appears that the evolution of Fis from an NtrC-like specific-binding transcription factor into a versatile nonspecific DNA-bending protein occurred with very subtle changes to the number and positioning of specific and nonspecific contacts.

Chapter 3: NtrC-like C-terminal DNA-binding domains: structure, function, and flexible tethering to the central domain

3.1 Summary

In this Chapter, I present the structures of the DNA-binding domains of NtrC1 and NtrC2, enabling an extensive cross-comparison of this domain with the homologous structures of the DNA-binding domains of NtrC4, NtrC, ZraR, and FIS. The N-terminal linkers that connect the DNA-binding domain to the central domain were studied in NtrC1 and NtrC2 and found to be unstructured. Additionally, the full-length NtrC1 crystal structure was solved, but the DNA-binding domains were disordered, providing further evidence that the C-D linker functions as a flexible tether.

3.2 Introduction

3.2.1 DNA-binding domains of σ^{54} activators

In *A. aeolicus* NtrCs as well as mesophilic NtrC, PspF, ZraR and DctD, the N-terminal R and C domains have been studied extensively (Batchelor 2008, Doucleff 2005, Lee 2003, Hastings 2003, Zhang, 2002). Less structural work has focused on the C-terminal DBDs and the linker regions that connect this domain to the central domains (CD-linker). Only two high-resolution DBD structures have been described in the literature: NtrC-DBD (Pelton 1999) and ZraR-C-DBD from *Salmonella enterica, serovar typhimurium* (Sallai 2005). Secondary structure and DNA-binding features of the DNA-binding domain of NifA were also discussed by Ray, *et al.*(2002). Chapter 2 also describes new structures of the 4DBD-free (Batchelor, 2008) and a 4DBD-DNA complex. These structures all demonstrate that the DBD exhibits a dimeric helix-turn-helix motif, a motif that is extremely common among transcription factors and other DNA-binding proteins (Aravind, 2009; Rohs 2010; Huffman and Brennan, 2002).

3.2.2 Functions of the DNA-binding domain

The function of the DBD in transcriptional activators has been studied extensively biochemically. It has been shown that one function of the NtrC DBD is to colocalize the EBP near the σ^{54} promoter, increasing the probability of an interaction with RNAP-holoenzyme (North et. al, 1997). A mutant form of NtrC from *Salmonella enterica serovar typhimurium* with three alanine substitutions that abolish DNA-binding was found to have 5000-fold reduced ability to activate transcription *in vitro*. However, wild-type activation rates could be reproduced with this mutant by raising the activator concentration 50-fold, although in this case the activators are no longer gene-specific

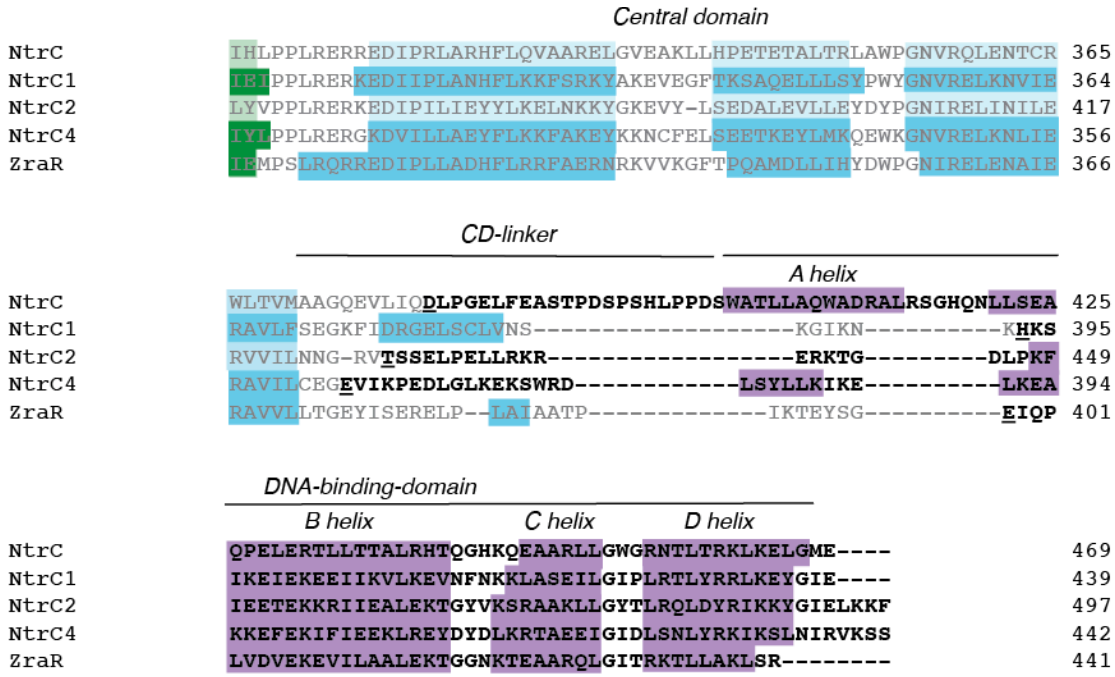
activators (Porter *et al.,* 1993). However, the proximity of the DNA-binding site to σ^{54} is also not vital, as it was shown that NtrC could still activate transcription when the NtrC binding sites were moved one kilobase upstream of their native binding site (Schulz 2000), or when attached to a separated, concatenated strand of circular DNA (Wedel 1990). Thus, the DBD appears to tether the activator near the σ^{54} activator, in order to raise the local concentration of activator with respect to the RNAP holoenzyme.
A second function of DNA-binding is the placement of two NtrC unactivated dimers next to one another, which may enhance the rapid oligomerization of the protein upon activation (Chen and Reitzer, 1995, Porter 1993). In fact, these two binding sites often exhibit cooperative binding, which raises the probability that both sites will be occupied at once. It has been shown that both NtrC and a homologous σ^{54}-activator DctD bind two tandem DNA sites cooperatively (Weiss et. al, 1992; Reitzer et al, 1989; Scholl and Nixon, 1996), and in the NorR system, three different cooperative binding sites are required for maximum activity (Tucker, et al., 2010). Since transcriptional activity requires the oligomerization of the activator protein, the two or three activator-binding sites would raise the local concentration of activator to stimulate this process, even if intracellular activator concentrations were low.

A third observed function of the DBD is the stabilization of the off-state dimer. In the case of NtrC, the DBD is the dimerization determinant for the full-length protein, and deletion of this domain produces monomeric protein with reduced ATPase activity (Klose et al, 1994). However, in NtrC1 and marginally in NtrC4, the unactivated RC domains without the DBDs are inherently dimeric (Batchelor, 2008; Doucleff 2005). It has been suggested that the presence of a strong dimerization interface in the RC domains is an indicator of positive or negative regulation of the ATPase domain (Doucleff 2005). For example, in the case of NtrC, activation is positively regulated: activation causes the monomeric R and C domains to structurally stabilize the formation of the inherently unstable hexameric ATPase ring (de Carlo, 2006). The DBD is the only domain that stabilizes the inactive dimer. In contrast, NtrC1 and to a weaker extent, NtrC4, exhibit negative regulation (Figure 3.4). The central domain is inherently able to form an active ATPase, but the RC dimerization interface represses the formation of the oligomer until the dimerization interface in the R domains is broken by activation (Doucleff, *et al.,* 2001). The size of this RC dimerization interface in NtrC1 and NtrC4 has also been implicated in tuning the degree of ATPase repression. In NtrC1, the RC domains form a very stable dimer with a coiled coil at the dimerization interface that strongly represses oligomerization. However, in NtrC4, the lack of a coiled coil linker at the RC dimer interface produces a less stable dimer that only weakly represses oligomerization (Batchelor, et al., 2008 and Doucleff, et al., 2001). These three examples constitute a spectrum of regulation: strong negative regulation (NtrC1), weak negative regulation (NtrC4), and positive regulation (NtrC), depending on the degree of dimerization of the RC domains.

3.2.3. CD-linker regions

In the previous discussions of regulatory mechanism and NtrC activation, the roles of DBDs and the CD-linker regions connecting the C and DBD domains have been less addressed. The role of the CD-linker is particularly interesting in light of the recent

Figure 3.1: Sequence alignment of the C-terminal and linker regions of NtrC, NtrC1, NtrC2, NtrC3, NtrC4, and ZraR. NtrC and ZraR sequences are from S. enterica serovar Typhimurium, all others are from A. aeolicus. Blue and purple regions mark α-helices, green regions mark β-sheets. Full-color regions are experimentally determined from structures; faded regions are inferred from secondary structure prediction and homology when no structural information was available. Boldface regions describe the constructs used in this work, or, in the case of ZraR, the region that was visible in the electron density.



study on full-length, oligomerized NtrC in the ADP-bound form by cryo-EM and the transition state AlFx-bound form by SAXS (de Carlo, 2006). Although DBD density was not visible in the ADP-bound state, ordered density, roughly correlated with the volume of three DNA-binding domains, appeared to dock under the hexameric ring in the ADP-AlF_x-bound form, the "transition state" analog. Since limited structural information is available about the CD-linker in any state, it is not yet possible to distinguish whether this docking process seen in NtrC is mediated by structure in the linker region, or if the linker serves as a flexible tether that brings the DBD in close proximity to the C domain. In other protein systems, there are numerous examples of linkers that act as simple tethers (immunoglobulins (Ababou 2008); leucyl-tRNA synthetase (Hsu 2008); Xylanase Cex (Poon 2007)), as structured mediators of docking interactions or cellular cross-talk (Smad3(Vasilaki 2009); DNA Pol (Klinge 2009); poly(rC) binding protein 2 (Sean 2008)), or as regions that can interconvert between structured and unstructured states in order to mediate docking/binding interactions (lac repressor (Swint-Kruse 2009); retinoblastoma protein (Burke 2010)).

To discuss the structure and thermodynamic dimerization stabilities of NtrC DBDs as well as the structure of the CD-linker regions, we present the structures of the DBDs of *A. aeolicus* NtrC1 and NtrC2, as well as information from a crystal structure of the full-length intact NtrC1. These structures, together with the structure of the NtrC4-

DBD (Chapter 2), *S. typhimurium* NtrC (Pelton, 1999) and ZraR (Sallai et al., 2005), and the highly homologous DNA-bending protein, the Factor for Inversion Stimulation (Fis), provide a family of seven homologous DNA-binding domains from diverse organisms that enables an extensive structural cross-comparison. We find that the size of the dimerization tendencies in NtrC, NtrC1, and NtrC4 DBDs is the inverse of the dimerization tendencies of each protein's RC domains. We also provide evidence that the CD-linker region of σ^{54} homologues functions as an unstructured, flexible tether.

3.3 Materials and Methods

3.3.1. Constructs

Secondary structure prediction (fig. 3.1) suggests that the poorly-conserved linker region between the conserved C and D domains in NtrC1, NtrC2, and NtrC4 is unstructured. Constructs were selected starting from a point in this flexible region and ending at the C-terminus of the full-length protein. Two different constructs of NtrC1DBD were made, 1DBD-long (373-439), including 20 residues of the N-terminal flexible linker, and 1DBD-short (393-439) that completely removes the N-terminal linker region. The construct of NtrC2 includes residues 428-497 (2DBD). The NtrC4 construct was made for residues 365-442 (4DBD). The constructs that were studied structurally, 1DBD-short, 2DBD, and 4DBD, are marked in bold in Figure 3.1.

3.3.2. Protein Expression and Purification

1DBD (residues 392-439) and 4DBD (residues 428-497) were cloned into the pet21a vector (Johanna Heideker, Dr. Ann Maris), and 2DBD (residues 365-442) was cloned into the pet21b vector (2DBD) as a $(His)_6$tag-fusion protein (Dr. Eunmi Hong). Cells were transformed into *E.coli* BL21(DE3) with a Rosetta.pLysS plasmid, and grown to an OD_{600} of 0.8 (1DBD, 4DBD) or 0.5 (2DBD) in Luria Broth media. Cells were induced with 1 mM isopropyl b-D-thiogalactopyranoside (IPTG) and harvested after 12-16 hours at 25°C (1DBD, 4DBD) and 4-6 hours at 37°C (2DBD). Isotopically labeled proteins were switched to ^{15}N or $^{15}N/^{13}C$-labeled M9 media (Cambridge Isotope Laboratory) before induction as described (Marley, 2001) and induced for 4-6 hours. Cells were harvested by centrifugation at 5000 rpm for 30 minutes. Cell pellets were suspended in lysis buffer (1DBD/4DBD: 50 mM sodium phosphate pH 6.5, 200 mM NaCl, 1 mM EDTA, and 0.1 mM PMSF; 2DBD: 20 mM sodium phosphate pH 7.0, 0.5 M NaCl, 0.1 mM PMSF), and disrupted by sonication. 1DBD and 4DBD were additionally heated to 80°C for 30 minutes to precipitate *E. coli* proteins. Lysates were then centrifuged at 30,000g for 30min, and supernatants were filtered using 0.45 μm filters to remove insoluble material. 1DBD and 4DBD were purified by Heparin affinity chromatography and eluted with a continuous linear NaCl gradient (200 mM-700 mM). DBD2 was loaded onto a HiTrap™Chelation HP (Pharmacia Biotech) column, and bound protein was eluted with a continuous 100 mL linear gradient of imidazole (0 - 500 mM). The His-tag was cleaved from 2DBD by tobacco etch virus (TEV) protease at 30°C for 6 hours. The cleaved His_6-tag and un-cleaved protein were removed by loading the digested

proteins over a HiTrap™Chelation HP. 1DBD, 2DBD, and 4DBD were finally purified by gel filtration on a HiLoad Superdex 75 prep grade column (Pharmacia) in gel filtration buffer (1DBD and 4DBD: 50 mM sodium phosphate buffer pH 6.5, 100 mM NaCl, 1 mM EDTA; 2DBD: 20 mM sodium phosphate pH 7.0, 100 mM NaCl). Samples were concentrated using Amicon Ultra centrifugal concentrators (5 kDa cutoff; Millipore).

3.3.3. Analytical ultracentrifugation

Analytical ultracentrifugation of 1DBD-long and 4DBD was carried out on a Beckman ultracentrifuge model XL-I using a 12-mm six-channel centerpiece and an An-50 Ti rotor. Data was collected at 20000 rpm at 280 nm and 300 nm for protein concentrations of 0.5, 1.0 and 1.4 mg/ml in 250 mM NaCl, 50 mM $NaPO_4$ pH 6.5. The run was allowed to reach equilibrium by spinning the samples for 24 hours. Velocity scans were taken in 10 replicates for each sample at 20°C. The radial step size was 0.001. Data analysis was performed using the software UltraScan-7.0.

3.3.4. Structural comparisons

Structures were analyzed with PYMOL. Helix angles, distances, and buried surface area were calculated with CHIMERA (Pettersen, et al., 2004).

3.4. Results

3.4.1. Construct selection of 1DBD and 2DBD

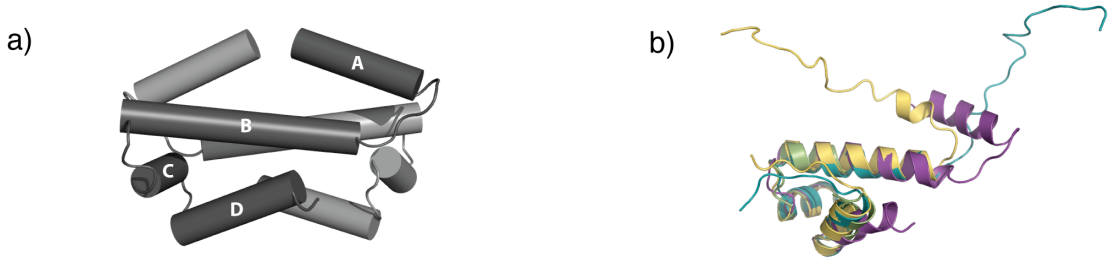
A multiple-sequence alignment with the C and DBD domains of NtrC, NtrC1, NtrC2, NtrC4, and ZraR proteins is shown in Figure 3.1. The constructs of NtrC1 and NtrC2, as well as NtrC4, NtrC, and the regions of visible DBD density in the ZraR crystal structure, are shown in boldface. Since the ZraR crystal structure does not contain an N-terminal A helix and the NtrC protein does, the nomenclature of this paper will count residues that align with residues in the A helix as belonging to the DNA-binding domain, even if these residues are disordered. The remaining residues between this A-helix region and the highly-conserved final helix in the central domain will be called the "CD-linker" region (Figure 3).

Since the N-terminal CD-linker and A-helix-DBD region was removed in the 1DBD construct used for crystallography (1DBD-short), a second construct, (1DBD-long) was prepared that included residues 373-439 for NMR analysis.

3.4.2. The structures of NtrC1-DBD and NtrC2-DBD

1DBD-short was crystallized and the structure was solved to 1.3 Å by Dr. Ann Maris and Anisa Young. The dimer crystallized in the asymmetric unit. The monomeric structure of 2DBD was solved by NMR spectroscopy in solution to 0.14 ± 0.05 Å backbone RMSD (for structured residues 448-492).

Figure 3.2. Structures of the NtrC DNA-binding domains. NtrC (purple), NtrC1(green), NtrC2(teal), NtrC4(yellow), Fis(dark red). a) Helix nomenclature in helix-turn-helix dimers. b) Overlay of monomeric subunits of NtrC (purple), NtrC1 (green), NtrC2 (teal), and NtrC4 (yellow) DBDs.

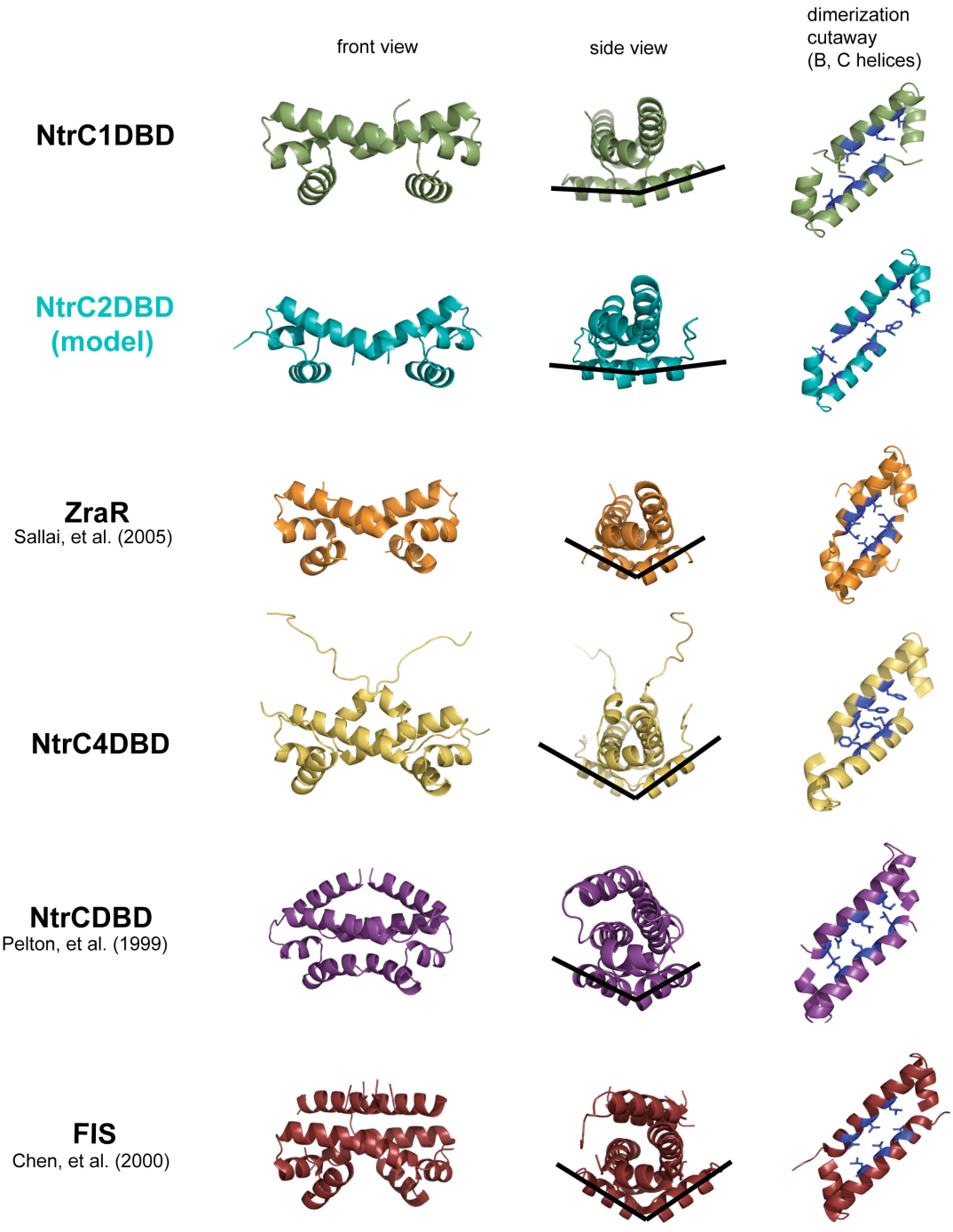


The structures of the solved DBDs and their homologues are shown in Figures 3.2 and 3.3. Helices (from the N terminus to the C terminus) are labeled A, B (dimerization helix), C (buttressing helix), and D (recognition helix), in accordance with naming conventions for helix-turn-helix proteins (Figure 1a) (Arivand, et al., 2005). The monomeric units of helices B-D, which make up the helix-turn-helix bundle, overlay closely in NtrC-DBD, 1DBD, 2DBD, and 4DBD, with an average RMSD of 1.42 Å (Fig. 3.2b). In this region, there are only slight differences between the structures: NtrC-DBD has a 37° kink in the D helix and a ~14° kink in the B helix, which is not present in the *A. aeolicus* NtrCs. The length of the B helix also varies, in 1DBD-short (16 residues), 2DBD (18 residues), 4DBD (20 residues), and NtrC-DBD (21 residues). 4DBD and NtrC-DBD have the additional N-terminal A helix that contributes to the dimerization interface. It is not present in the monomeric NtrC2 structure, and its presence is unknown in 1DBD-short because the short construct removed N-terminal residues that might be involved in its formation.

More dramatic differences between structures are immediately visible in the dimeric structures of NtrC-DBD, 1DBD, and 4DBD. This is mostly due to the dimerization angle of the B helices, which produces different hinge-motion-related positions of the recognition helices, resulting in an average 2.70 Å RMSD. In NtrC-DBD, 1DBD, and 4DBD, the angles formed by the B helices are 14°, 33°, and 42°, and distances between the central axis of the recognition helices are 18 Å, 22 Å, and 28 Å, respectively, calculated between central helical axes. This suggests that slight variation in the hydrophobic packing of the dimerization interface can result in alteration of the length of the DNA element recognized by the DNA-binding domain. This also causes variation in the angles formed by the D helices in contacting DNA (Figure 3.3, side view).

Details regarding the DBD structures and sequences are summarized in Table 3.1. Figure 3.3 presents a number of views of the DBDs from NtrC, NtrC1, NtrC2, and NtrC4 with the DBDs of ZraR and FIS for structural comparison.

Figure 3.3. Comparison of DNA-binding domains from A. aeolicus and other homologues. The NtrC2-DBD is modeled as a dimer from the monomer structure with HADDOCK.



3.4.3. Dimerization interfaces: 1DBD and 2DBD are monomeric in solution

To understand the thermodynamic dimerization stabilities of NtrC-like proteins, we performed AUC (analytical ultracentrifugation) on 1DBD-short and 4DBD. Although 1DBD-short crystallized as a dimer, it was found to be monomeric in solution. To ensure that this weak dimerization was not due to the truncation of the N-terminal linker, 1DBD-long was prepared and also found to be monomeric with AUC. To supplement this finding, both constructs of 1DBD were also analyzed by mass spectrometry, and confirmed to be almost exclusively monomeric, suggesting that the N-terminal linker contains negligible dimeric stabilization. 2DBD, which already includes 18 residues of the CD-linker, was also found to be monomeric in solution by NMR (Dr. Eunmi Hong). 4DBD, however, was found to be dimeric by both AUC and mass spectrometry (Johanna Heideker; Harry Sterling).

The dimerization interfaces in each of these proteins explain these variations (Figure 2). The intermolecular interfaces in all structures are primarily formed by the residues in helix B (NtrC1-DBD: 396-411, NtrC2-DBD: 448-465, NtrC4-DBD: 402-421, NtrC-DBD: 421-441) and from contact interfaces on helix A (in the case of NtrC4 and NtrC). In NtrC-DBD, dimerization interactions consist of the intercalation of three leucine residues (L422, L429, L433) on each B helix, and hydrophobic packing of residues on the A helix against the other monomer's B helix, forming a long hydrophobic core that buries ~2,600 $Å^2$ of solvent-accessible surface area. In 4DBD, the dimerization interface consists of the aromatic ring-stacking interactions of two phenylalanine residues (F398, F402) on each helix as well as a Leu382-Leu382 interaction on the A helices that together bury ~2,400 $Å^2$ of surface area. In NtrC1, the dimerization interface is notably smaller: its main determinant is an Ile399-Ile399 interaction on the crossed B helices, and a limited number of van der Waals interactions at the base of the three-helix bundle junction that bury only ~950 $Å^2$. To investigate possible interactions at the dimeric interface of monomeric 2DBD, the structure of the symmetric dimer was modeled by docking two monomers together with the protein-docking program HADDOCK (Dominguez, et al., 2003, and Vries et al., 2007). The lowest energy structure of this NtrC2-DBD dimeric model (with a HADDOCK score of –97 ± 7) overlays closely with the NtrC1-DBD structure with an RMSD of 2.92 Å. The only hydrophobic residues on the B helix that are positioned in the dimerization interface region are F449, I450, and T453. In this model, the I450 residues on opposing B helices, as in NtrC1-DBD, form a van der Waals contact, the two T453 residues form hydrogen bonds to the carbonyls in the backbone of each F449, and additional inter-monomeric hydrophobic contacts are formed between the F449 side chains with the LGY sequence at the three-helix bundle junction, burying a total of ~1000 $Å^2$.

3.4.4. Crystal structure of full-length NtrC1: the DBD is flexibly tethered to the C domain

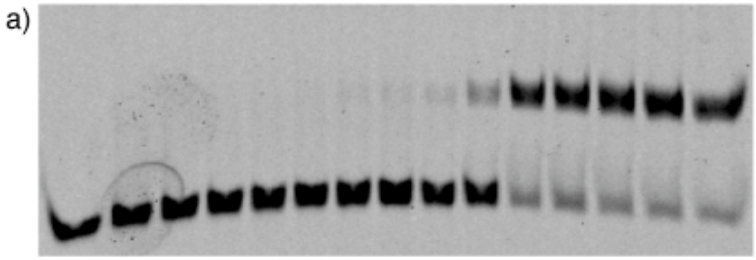
Currently, no structure has been reported of a full-length intact σ^{54} activator, possibly due to the difficulty crystallizing multidomain proteins with flexible linkers. However, a full-length construct of NtrC1 was crystallized and the structure was solved to 2.2 Å resolution in our lab by Dr. Ann Maris. Unfortunately, this structure shows a

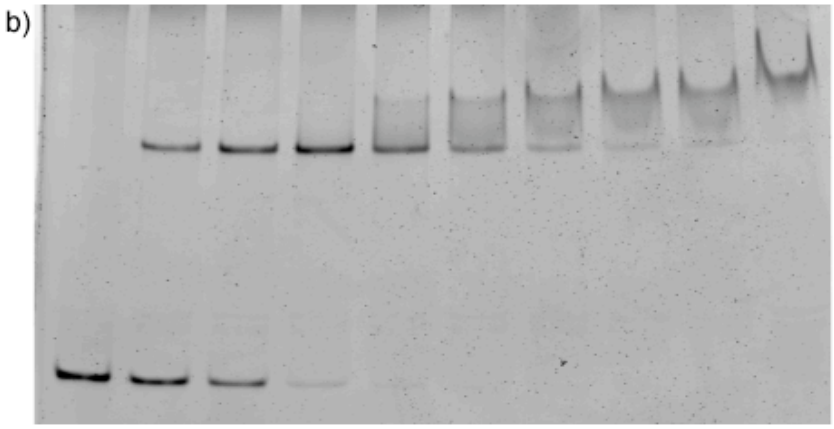
number of non-native features, including a L183S mutation and a disulfide linkage between loop residues Cys158 and Cys161. However, the most interesting feature of this structure was that almost no electron density was visible in the region where the DNA-binding domains are expected, suggesting that the DNA-binding domains' position is disordered in the crystal. A gap between NtrC1-RC dimers, large enough to accommodate the DNA-binding domains, separates the central domains of one dimer from the receiver domains of its symmetry mate (see Figure 4). The structure of NtrC1-RC in the unactivated state was solved previously (Lee *et al*.2003). Therefore, the main value of the full-length NtrC1 structure is the observation that only very weak electron density can be observed in the region where the DBD must be, indicating that its position within the crystal is variable. This provides additional evidence that the DBD is flexibly linked to the central domain. It also suggests that there is not a significant docking interface between the central and DNA-binding domains in the unactivated state. The vertical distance between the central domains of one NtrC1-RC dimer and the molecule immediately adjacent in the crystal matches the height of the NtrC1-DBD (26 Å), constraining its vertical position. However, the DBD is less constrained in the horizontal dimension. The sparse lateral contacts enable the DNA-binding domain to have rotational freedom around the vertical axis as well as some translational freedom (Figure 3.5), explaining the lack of electron density.

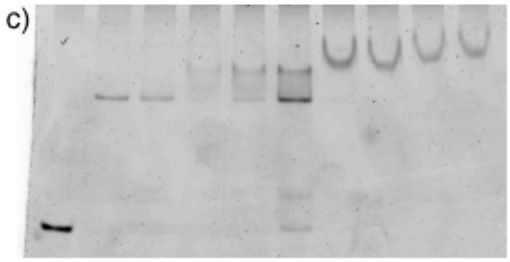
3.4.5. Cooperativity of binding in NtrC4-FL

NtrC has been shown to bind cooperatively to its high- and low-affinity binding sites. Whether NtrC4 also binds DNA cooperatively is not known. To probe for cooperative binding, the binding of full-length NtrC4 to the 21-mer low-affinity binding site, and to 100-base pair UAS regions that include the high and low-affinity sites, was analyzed by gel-shift assays (Figure 3.4). Gel-shifts on the native low-affinity binding site demonstrate no binding at low protein concentration, suggesting non-two-state binding or cooperativity, as seen in Chapter 2 (figure 3.4a). The K_d of binding is estimated at ~ 200 nM, which is very weak compared to the high-affinity binding site (K_d = 1.3 nM). To determine whether DNA-binding is enhanced at the low-affinity site when the high-affinity site is occupied, a 100-base pair region of the *lpxC* UAS was amplified by PCR, including both the high-affinity and low-affinity binding sites. It was found that binding at the low-affinity site is weakly enhanced by the presence of the high-affinity site. Although the doubly-occupied site produces a band that is smeary and not suitable for quantitative K_d analysis (Figure 3.4b), it appears to half-deplete the singly-bound site at ~ 70 nM, which would correspond to a ~3-fold binding enhancement. At higher protein concentration, smiling bands suggest extensive protein aggregation, due to nonspecific aggregation or oligomerization (Figure 3.4c, d).

Figure 3.4. Gel-shift assays with NtrC4-FL andbinding regions including the low-affinity binding site. a) Gel-shift assay with the fluorscein-labeled 21-mer low-affinity binding site. DNA concentration is 25 nM; protein dimer concentrations (from left to right) are 0, 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 125, 187.5, 250, 312.5, 375, 437.5, and 500 nM. b) Binding to the 100-mer native binding site including high and low affinity binding sites. DNA concentration is 10 nM; protein dimer concentrations are 0, 10, 20, 30, 50, 70, 90, 110, 130, and 200 nM. Comparison of binding to the c) native 100-mer site or d) 100-mer site with the low-affinity binding site removed; DNA concentration is 10 nM; protein dimer concentrations are 0, 10, 20, 40, 70, 100, 150, 200, 250, and 300 nM.







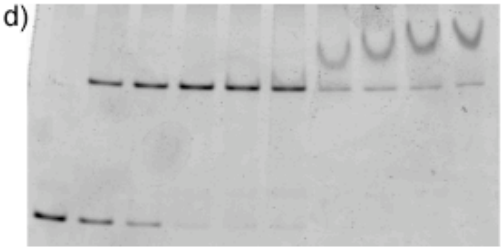
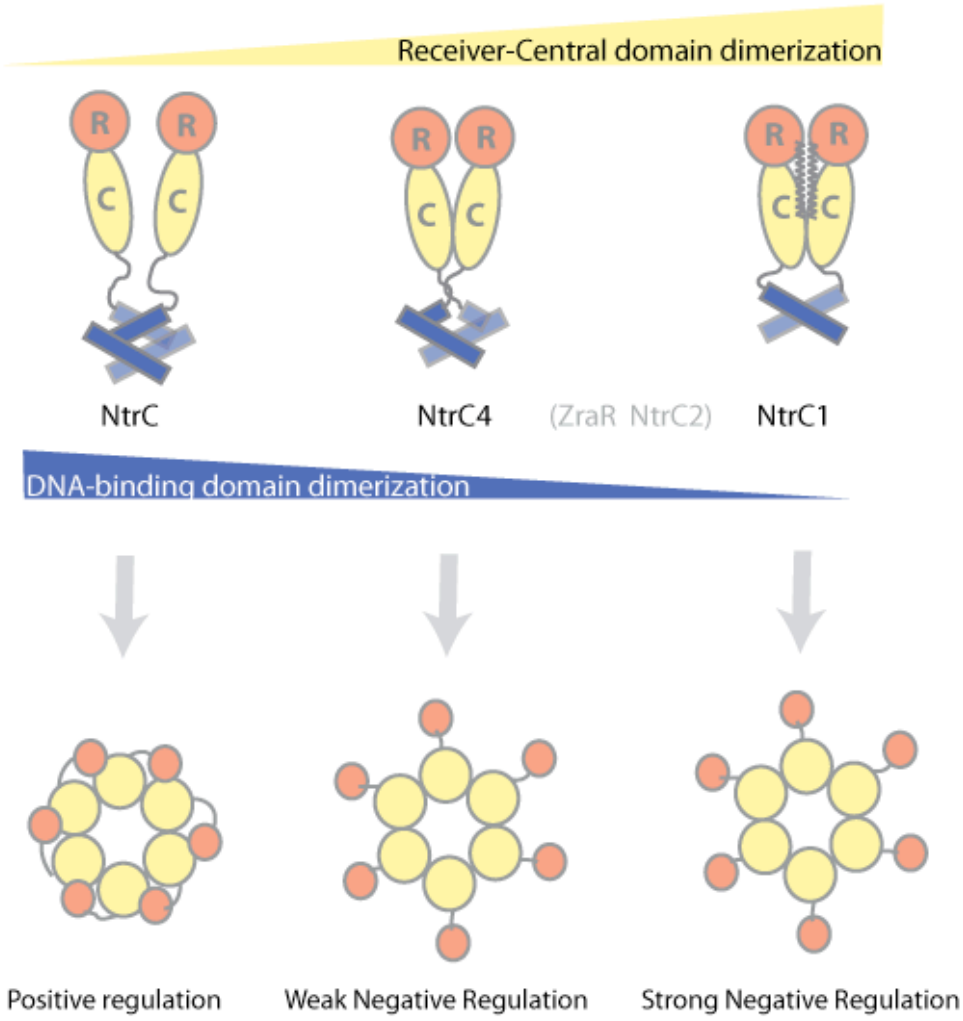


Figure 3.5. Summary of domain architecture and regulatory mechanism in NtrC, NtrC4, and NtrC1. Based on this work, ZraR and NtrC2 would be predicted to exhibit a mechanism between NtrC4 and NtrC1 regulation.



3.5. Discussion

3.5.1. DBD and RC dimer stability complement one another

In the DNA-binding domains of NtrC, NtrC4, NtrC2 (model), and NtrC1, the buried surface area due to dimerization is 2,600 Å, 2,390 Å, 1080 Å, and 1050 Å, respectively. In contrast, the size of the dimerization interface of the RC domains

increases in the order: NtrC, NtrC4, NtrC1 (Figure 3.5). The stability or even existence of an NtrC2-RC dimer is not known, but secondary structure prediction suggests that NtrC2 also contains the coiled-coil helix that causes stable dimerization of the NtrC1-RC domains. The fact that the buried surface area due to DBD dimerization is inversely proportional to the stability of the RC dimer suggests that the DNA-binding domains fulfill residual dimerization needs of the protein. Currently, there are no known NtrC-like proteins that are not dimeric when unactivated, underscoring the importance of dimerization. If dimerization is provided by the RC domains, more evolutionary freedom may be available to tune the DNA-binding domain function or complex interface without destabilizing dimerization. The inverse is also true; dimerization at the DBD would provide more evolutionary freedom for variation at the RC domains.

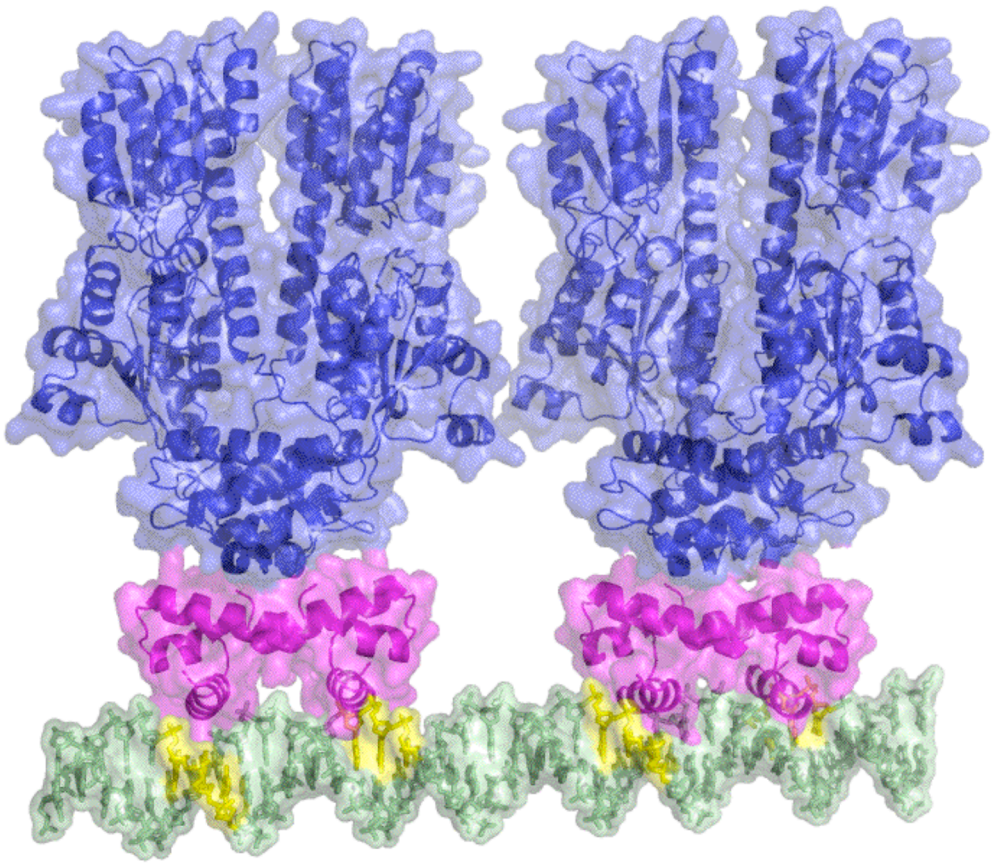
The presence of the fourth, "A" helix in some NtrC DBDs appears to be a strong determinant of dimer stability. In the case of the NtrC-DBD homologue, Fis, which was discussed extensively in Chapter 2, the extended dimerization interface formed by the long A helix stabilizes the dimer (Cheng 2000). NtrC's DNA-binding domain also has an A helix, and forms a very stable dimer (Pelton, et al., 1999). Since Fis and NtrC are both shown/predicted to bend DNA, it is possible that the A helix, or some other form of dimer stabilization, may be required for DNA-bending. Consistent with this hypothesis, NtrC4-DBD, which has an abbreviated A helix, was shown to cause a very slight bend in the DNA upon binding (Chapter 2; Cheng 2000).

For NtrC1 and NtrC2 DNA-binding domains, secondary structure prediction, NMR, and mass spectrometry data suggest that the N-terminal region does not contain a stable A helix, at least in the absence of DNA. This indicates that NtrC1 and NtrC2 share more similarity with ZraR, which apparently does not have an A helix in the DBD (Sallai and Tucker, 2005). The structures of the ZraR-DBD and the NtrC1-DBD overlay closely with an RMSD of 2.2 Å, have similar spacing between the recognition helices (ZraR: 16 Å, 1DBD: ~18Å), and bury similar solvent-accessible surface areas (ZraR: ~1070 $Å^2$, 1DBD: ~1050$Å^2$). One notable difference between 1DBD and ZraR is that ZraR has two hydrophobic contacts between the B helices (V405 and V409), whereas NtrC1-DBD has only one (I399). This would be expected to make the ZraR dimer marginally more stable than the 1DBD dimer.

However, although NtrC1 exhibits no predicted or observed structure in the A-helix DBD region, an additional C-terminal helix was observed to pack against the C-terminal helices in central domain in the NtrC1-RC crystal structure, in the CD-linker region (Figure 3.1) (Lee, et al., 2003). One can speculate that this helix compensates for the lack of the A helix in NtrC1. The overall length of the NtrC1 CD-linker is shorter than in NtrC, NtrC2, and NtrC4, thus, the presence of the additional C-domain helix may act to shorten the tether between the C and DBD domains, and stabilize the weaker dimerization interface of the DNA-binding domain. This may also reduce the flexibility between the C and DBD domains, and assist the close-packing of two NtrC1 full-length dimers onto the tandem DNA-binding sites (Figure 3.6).

Although little is known about the structure of the unactivated ZraR R and C domains, it has been suggested that ZraR's DNA-binding domain may act as the dimerization determinant for the full-length protein, as seen in NtrC (Sallai, 2005). However, a number of features of the ZraR hexameric C-DBD crystal structure suggest that it bears more similarity to NtrC4 in terms of structure and regulatory mechanism.

Figure 3.6. A model of two dimeric NtrC1-full length activators on Aq_1119 native DNA tandem binding sites.



According to the coiled coil prediction program COILS (Lupas, 1991), ZraR appears to lack a coiled coil at the RC dimerization interface, suggesting that its dimerization interface is weak relative to NtrC1. The observation that the C-DBD domains oligomerized into a hexamer in the crystals suggests that it is negatively regulated by its R domains. The structural similarity of ZraR's DNA-binding domain to the predominantly-monomeric 1DBD suggests that ZraR would not successfully dimerize if it relied on the DNA-binding domains alone.

In 1DBD, it is possible that the Ile399 hydrophobic contact constitutes a pivot point, allowing flexibility in the dimerization interface that enables the recognition helices to swing outward. Pivots with functional purposes have been documented in helix-turn-helix DNA-binding proteins in the cases of the LacI headpiece (Kalodimos et al., 2001) and the phage-encoded activator protein Mor (Kumaraswami et al., 2004),

where the pivot enables the protein to switch between specific and nonspecific interaction modes, or recognize multiple binding sites. The role for a swivel in NtrC1 is not apparent at this point.

3.5.2. DNA-binding and bending

It is well-documented that Fis can bend DNA from 40-90°, depending on the DNA sequence, and that this bending is important for activation of transcription at the ribosomal RNA promoter rrnB P1 (Cheng et al., 2000). Fis, like NtrC, has recognition helices that are spaced too closely to interact with B-form DNA without distorting it. However, as is observed in 1DBD and 4DBD and modeled in 2DBD, the wide spacing between the recognition helices suggests that DNA-bending is not necessary for interaction with DNA for these proteins. This may be due to an evolutionary trend that eliminated the need for DNA bending in certain organisms, or it may be a reflection of the thermophilic nature of NtrC proteins in *A. aeolicus*.

3.5.3. Absence of prolines in *A. aeolicus* NtrCs

It has been previously discussed that NtrC and Fis both contain a proline residue at one of two positions in the B helix, producing a slight kink, that is conserved in a number of α and γ-proteobacterial species (Pelton, 1999; Morett & Bork, 1998). In the case of Fis, this kink has been shown to be necessary for Hin-mediated DNA inversion activity (Osuna et al., 1991), and the crystal structure of a Pro61Ala mutant was found to exhibit the kinked helix in spite of lacking the proline, due to hydrophobic and hydrogen bond interactions that stabilize the kinked form (Yuan et al., 1994). It is interesting, therefore, that although the proline residue is conserved in NtrC, it is not present in any of the *A. aeolicus* proteins presented here. A recent sequence alignment of NtrC4 with the top 500 hits in the Blast non-redundant database found only one hit in which the Proline was found at a residue predicted to be located anywhere in the B helix. This suggests that the proline residue in NtrC and Fis proteins is not needed for function in *A. aeolicus* forms of the protein.

3.5.4. NtrC DBDs and cooperative binding

NtrC has been shown to bind cooperatively to the two tandem binding sites upstream from the glnA promoter (Porter 1993). It is unlikely that this cooperativity originates from contacts between DBDs of consecutive dimers, since the tandem binding sites are separated by ~17 base pairs or ~60 Å. However, this cooperativity could be mediated by the R and/or C domains. Unpublished data shows that the full-length NtrC1 also binds cooperatively (Tracy Nixon, personal communication). In this case, the distance between high affinity binding sites is only 9 base pairs, or ~30 Å, apart. Placing rigid models of two full-length NtrC1 dimers on ideal B-form DNA suggests that intermolecular contacts between the dimers would be unavoidable (Figure 3.6). To explore whether cooperativity occurs in the binding of NtrC4 to its tandem binding sites, gel shift assays were performed. Interestingly, although NtrC4 would not be predicted to bind cooperatively based on the long distance between binding sites (sites are separated

by ~40 base pairs, or ~135 Å), gel shifts indicate that binding to the second site exhibits a weak, but clear enhancement if the first binding site is present (Figure 3.4). The height and width of NtrC4-RC dimers (68 and 83 Å, respectively) suggest that cooperativity would not be possible if the proteins docked as rigid bodies onto straight, B-form DNA. Therefore, cooperativity either depends upon distortion of the DNA, such as DNA-looping or curvature, or due to the flexibility conferred by the long CD-linkers present in NtrC4. Since the NtrC4 CD-linker is the longest of the *A. aeolicus* CD-linker regions, the NtrC4-RC domains may have longer tethers to the DBD, enabling them to interact over a longer distance.

3.5.5. CD linkers of σ^{54} activators are disordered

Although the CD-linker regions were thought to be largely disordered based upon sequence analysis, there had been little direct study of these regions. Secondary structure prediction (Jpred) does not predict any secondary structure in the CD-linker regions (Figure 3) for *Aquifex aeolicus* NtrC's or mesophilic NtrC. Additionally, the NMR structure of NtrC from Salmonella typhimurium (Pelton, 1999) and the crystal structure of NtrC4-DBD from *A. aeolicus* (Chapter 2) have disordered N-terminal linkers. The only crystal structure of a σ^{54} activator including both the central domain and the DBD is the structure of ZraR, in which the hexameric central domain is clearly resolved, the DNA-binding domains formed dimers with crystallographic symmetry mates, and the CD linker is invisible (Sallai, et al., 2005).

This work provides additional evidence that the CD-linker in σ^{54}-proteins acts as an unstructured, flexible tether. The absence of DNA-binding domain density in the NtrC1-full length crystals as well as the lack of N-terminal structure in the ZraR, NtrC, NtrC1, NtrC2, and NtrC4 DNA-binding domains support this view. In the case of CD linker regions in NtrC, and ZraR, a high percentage of proline residues (16% and 10%) in the linker further reduce the likelihood of extended secondary structure in any activation state. Although it is not possible to rule out the possibility that regions of this linker acquire structure in the activated state, we would predict that DBD-docking onto the activated ATPase ring, as seen in the case of NtrC (de Carlo, 2006), is due to surface interactions that occur between the oligomerized C domains and the DBD, rather than being mediated by the CD-linker.

3.6. Future studies

3.6.1. NMR studies on the N-terminal linker of 1DBD-long

Although AUC, mass spec, and full-length NtrC1 crystallography all suggest that the N-terminal linker region of NtrC1 does not provide significant stabilization of the dimer of 1DBD, direct structural evidence is not available. One method to investigate this region structurally is by NMR. By assigning the backbone resonances in the HSQC spectrum of 1DBD-long, then searching for indicators of secondary structure in the N-terminal linker region (chemical shift indeces of the Cα and Hα shifts, and searching for α-helical NOEs), it would be possible to determine whether an A helix, or the C-terminal helix in the central domain, is present in this construct.

Chapter 4: NMR studies on region σ^N3 of σ^{54}

4.1 Summary

In this Chapter, I describe my work on region σ^N3 of the σ^{54} factor. I use TROSY-NMR to study the σ^N2-σ^N4 regions (residues 69-398) in free and DNA-bound forms. Although the linewidths in the TROSY spectra were sharp, only half of the peaks were visible, suggesting that they derive from the σ^N2 and σ^N4 regions only. A sample with specific $^1H/^{15}N$-Ile/Phe labels in a $^2H/^{14}N$ background was prepared, but the data quality suggests that protein aggregation may be occurring. Further optimization of the sample will be necessary to produce conclusive results.

4.2 Introduction

4.2.1 Structural studies on the σ^N3 region of the σ^{54} factor: current understanding

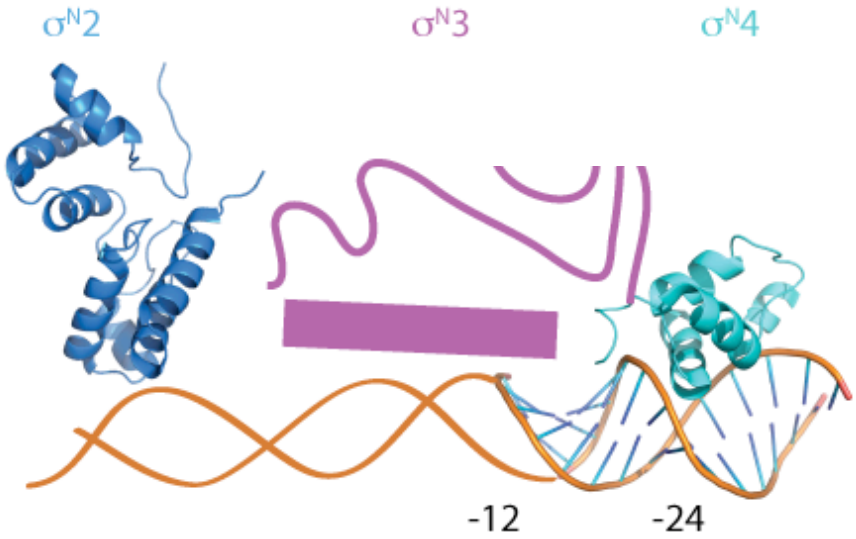
In the introductory chapter of this thesis, basic structural features of the σ^{54} factor were presented, including the description of regions σ^N1-σ^N4. Unlike the σ^{70} factor, a high-resolution structure of the full-length σ^{54} factor has not been determined. However, structures of the core-binding domain (region σ^N2) and the DNA-binding domain (region σ^N4), both free and in complex with the –24 site DNA are available (Hong, et al, 2008; Doucleff, et al, 2005; Doucleff et al., 2007). There is great interest in elucidating structure in the region between the core-binding domain and the DNA-binding domain, which I call σ^N3 (residues 200-331). This region is expected to include the –12-binding motif, as well as the residues that participate in DNA-melting. Currently, the conformational change that σ^{54} undergoes upon contacting the activator protein, and how this change transduces energy from ATP, is poorly understood. Although significant mechanistic information has come from cryo-EM studies on σ^{54}-RNAP (Section 1.4), understanding the full mechanism will require atomic-resolution information on the σ^N1 and σ^N3 regions.

4.2.2 X-ray crystallography of the σ^{54} factor

In the past decade, a number of laboratories have attempted to crystallize the σ^{54} factor, both alone and in complex with RNAP and/or DNA (personal communication). One promising effort comes from the Darst laboratory, which has crystallized and partially solved the σ^{54} factor regions σ^N2-σ^N4 in complex with DNA. The resolution of their data is 3.5 Å, however, the data quality has so far proven insufficient to build a complete structural model. Sidechains are generally not visible in the density, and there are regions of missing density in the σ^N3 region.

Figure 1. a) Schematic of structure-in-progress from the Darst lab: σ^{54}, 69-398 in complex with DNA. b) Sequence of the 69-398 construct with Jpred secondary structure predictions. Color scheme is the same in a) and b).

a)



b)

Core-binding domain (region σ^N2)

```
69  YTPSELEELQQNIKLELEGKEQELALELLNYLNEKGFLSKSVEEISDVLR
    ----HHHHHHHHH-----HHHHHHHHHHHH-----------HHHHHHH--

119 CSVEELEKVRQKVLRLEPLGVCSKDVWEFLELQIEEIYPEEEEILKKALR
    --HHHHHHHHHHH-----------HHHHHHHHHHH----HHHHHHHHHH

169 DLKRGKKLKPEIKGKLSRLRLFPLSSSAEKVYTFAKVDAIIEEENGEFFI
    HHHHHHHHHHHHHHHHHH------------EEEEEEEEE---EEEE
```

Middle region (region σ^N3)

```
219 YLYEDFIDIDLNEEYWELYKKSRNLQKELKEAFERYESIRKVLDIRRRNL
    EE-----EEEE-HHHHHHH----HHHHHHHHHHHHHHHHHHHHHHHHHH

269 RKVLEKIVERQKDFLTGKGSLKPLTLREVSSEIGIHESTLSRIVNSKYVK
    HHHHHHHHHHHH-----------HHHHHHHHH-----HHHHHH---EEE
```

DNA-binding RPON domain (region σ^N4)

```
319  TPVGTYSLRTFFVRESAEGLTQGELMKLIKEIVENEDKRKPYSDQEIANI
     ----EEEEE----------HHHHHHHHHHHHHHH---------HHHHHHH

369 LKEKGFKVARRTVAKYREMLGIPSSRERRI
    HHH----EEHHHHHHHHHHHH----------
```

Even though the NMR structures of the σ^N2 and σ^N4 domains could be modeled into the electron density, region σ^N3 cannot be traced unambiguously. It is possible that additional structural information on region σ^N3 may allow us to overcome current difficulties with structure building. A schematic of the current structural model obtained by the Darst lab is shown in Figure 1.

Solution NMR not only provides an excellent complement to X-ray crystallography, but also allows the characterization of protein dynamics, which are usually inaccessible to solid-state techniques. Additionally, a wealth of isotope labeling techniques can be exploited to highlight specific atoms or residues of interest and to facilitate data analysis. To assist the Darst lab in solving the crystal structure of σ^{54} I used NMR spectroscopy to explore structure in region σ^N3. Since this region was found to be relatively unstable in isolation, these studies were performed predominantly on a construct including the neighboring regions σ^N2-σ^N4, residues 69-398, a 38-kDa protein.

4.2.3 Introduction to NMR

Since the principles of NMR are the subject of a number of excellent books (Keeler, 2005, Wüthrich, 1986, Cavanagh, 1996) and beyond the scope of this thesis, only a simple treatment of the basic tenets relevant to my project are presented here.

4.2.3.1 NMR Basics

NMR spectroscopy takes advantage of a quantum effect existing in nuclear spins, especially those that are "spin 1/2" systems, including ^1H, ^{15}N, ^{13}C, and ^{31}P. The nuclei of these atoms have a slight inherent magnetization associated with angular momentum, that is known as spin. When these nuclei are placed in a magnetic field, they can exist in one of two states: a low-energy state aligned with the magnetic field, or a high-energy state antiparallel to the magnetic field. The distribution of spins aligned with the magnet is governed by the Boltzmann distribution, creating a slight overall magnetization of the sample. To simplify the discussion, we often speak of the overall magnetization of the entire sample, graphically represented by the vector M, instead of individual spins. The 3-dimensional direction and magnitude of this M vector can be manipulated by interaction with appropriately-tuned radiofrequency pulses. These pulses can be calibrated to tip the magnetization vector from its equilibrium position along the z axis into the transverse (x,y) plane. This tipping, as described by classical physics, produces a magnetization vector that precesses around the z axis as it returns to its equilibrium position. The frequency with which the spin precesses is known as the Larmor frequency, and is specific to the identity of the spin, as well as the spin's local environment. The behavior of this magnetization vector forms the basis of all NMR experiments. The full description of this process requires a quantum mechanical treatment.

4.2.3.2 T1, T2 relaxation processes

After spins are tipped off its axis by a radiofrequency pulse, it slowly returns to its equilibrium position in a phenomenon known as relaxation. There are two forms of

relaxation, called T1 and T2 relaxation. T1 relaxation, also known as longitudinal relaxation, describes the exponential process by which a spin returns to its equilibrium state along the z axis. T2 relaxation, also known as transverse relaxation, describes the dissipation of magnetization along the x, y axes. Since the focus of this thesis is T2 relaxation, and specifically T2 relaxation as it relates to the molecular size of a protein, the following discussion will focus exclusively on T2.

There are a number of sources of local magnetic fields, also known as relaxation mechanisms, that drive T2 relaxation. But for spin-1/2 systems, two relaxation mechanisms predominate: dipolar relaxation, and chemical shift anisotropy.

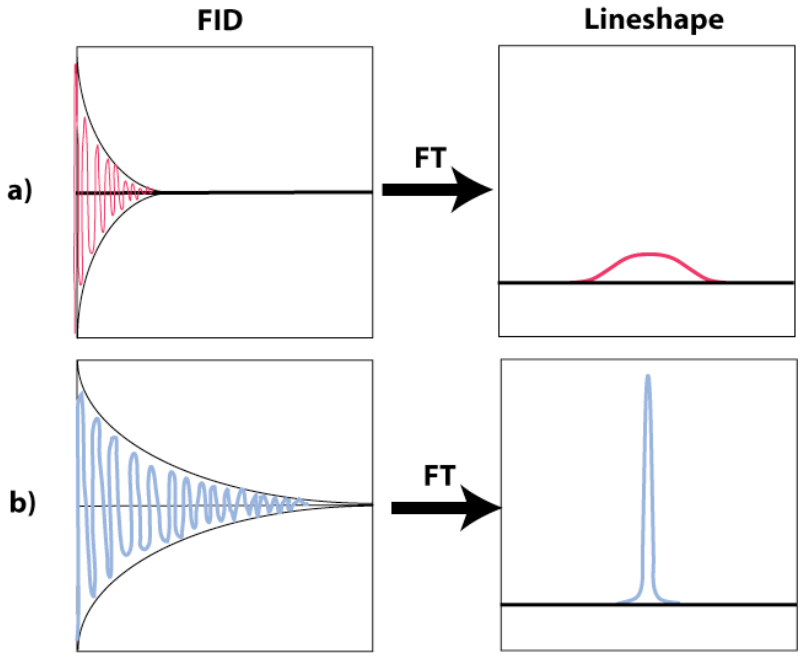
4.2.3.3 Relaxation mechanisms: dipolar relaxation and chemical shift anisotropy (CSA)

Dipolar relaxation derives from a local magnetic field produced by the magnetic moment of another spin. This two-spin relaxation phenomenon depends on three main parameters: the distance r between the two spins (it falls off as $1/r^6$), the gyromagnetic ratio (γ) of the spin generating the field, and the orientation of the vector joining the two spins relative to the applied magnetic field. The sharp distance dependence makes this effect negligible at distances larger than ~5Å. Since ^{1}H have a proportionally larger γ than ^{13}C (4x) or ^{15}N (10x), ^{1}H spins cause the majority of dipolar relaxation in proteins. Dipolar relaxation becomes more pronounced in systems that tumble slowly.
Chemical shift anisotropy (CSA), unlike dipolar relaxation, is an effect that derives from the local fields generated by electrons in a molecule. According to the laws of electromagnetism, bonding electrons placed in a strong applied field induce small, local fields. Thus, a nucleus experiences the sum of the applied and local fields, which shifts the Larmor frequency, causing a chemical shift. The amount by which the Larmor frequency is shifted depends upon the extent to which the local magnetic field varies with orientation. This, in turn, depends in part on the bonds present in a molecule. Since some spins, such as those in a C=O bond, may experience a higher orientation-dependent field than others, this generates anisotropy in the overall field perceived by the spins in a molecule. Like dipolar coupling, CSA depends on γ, and becomes more pronounced in systems that tumble slowly.

Although T2 relaxation can be directly utilized for a number of techniques in protein NMR spectroscopy, including diffusion measurements (Knauss, et al., 1996), CPMG experiments detecting chemical exchange (Dittmer and Bodenhausen, 2004), and paramagnetic relaxation enhancement (Clore, 2008), elevated rates of T2 relaxation often reduce the quality of NMR spectra. This is due to the inverse proportionality of the T2 relaxation time constant and linewidth, as shown in figure 2.

The line-broadening due to T2 relaxation is generally manageable in small proteins (MW< 30 kDa); however, it increases significantly in larger proteins (MW>30 kDa). This increases the incidence of peak-overlap and reduces the intensity of peaks,

Figure 2: The inverse proportionality of T2 and linewidths. Linewidths are proportional to molecular weight; for systems with molecular weights ≥ 30 kDa, some experiments produce poor data quality. Example FIDs and corresponding linewidths of a) a large and b) small protein.



which can compromise data quality significantly.

4.2.3.4 TROSY-NMR

The simultaneous presence of both dipolar relaxation and CSA interactions has an interesting side-effect that can be exploited for NMR of larger systems. As described above, both dipolar relaxation and CSA arise from the generation of random local fields that are time-dependent. For these two relaxation mechanisms to occur independently of one another, the sources of their time-dependence must be unrelated. However, in the case of large proteins, this is not the case. The correlation between the time-dependence of dipolar relaxation and CSA produces an affect known as *cross-correlation* that causes a mixing of relaxation mechanisms. Basic 1D NMR demonstrates that in a two-spin system, such as a ^{1}H-^{15}N amide bond in a protein backbone, a coupled ^{1}H NMR spectrum will show a doublet for the ^{1}H spin, reflecting the two different spin states for the ^{15}N atom (figure 3). Theory predicts that cross-correlation causes each peak in this doublet to relax at a different rate; in one of the doublets, the effects of dipolar relaxation and CSA are additive, but in the other, they are subtractive. This subtractive peak experiences

reduced T2 relaxation, and therefore has a sharper linewidth. In a 2-dimensional non-decoupled ^{1}H-^{15}N NMR spectrum, such as an HSQC (Heteronuclear Single Quantum Correlation Spectroscopy), a single ^{1}H-^{15}N bound pair will produce four peaks, doublets in each dimension, with varying linewidths (Figure 3b). This effect is at the heart of a technique called TROSY (Transverse Relaxation Optimized SpectroscopY). TROSY spectroscopy enables the isolation of the sharpest peak. This enables the collection of NMR spectra on much larger proteins (30 kDa < MW ~< 300 kDa) with lineshapes that are comparable to those of smaller systems. An example of non-TROSY and TROSY spectra from a 45 kDa protein (figure 3d) demonstrates the increase in data quality.

TROSY is also considerably more effective when combined with deuteration in proteins. Since the γ of ^{2}H is six times smaller than the γ of ^{1}H, replacing ^{1}H atoms in a protein with ^{2}H causes a large reduction in dipolar coupling. TROSY alone without deuteration is estimated to give a 60% reduction in ^{1}H linewidths, and a 40% reduction in ^{15}N linewidths. However, combined with deuteration, TROSY reduces the ^{1}H linewidths by 40-fold, and the ^{15}N linewidths 10-fold (Pervushin, et al., 1997). However, deuteration has a few experimental drawbacks. First, it is ~4x more expensive to deuterate proteins relative to standard $^{15}N/^{13}C$-labeling, and protein expression in D_2O media generally exhibits severely reduced yields. Second, removing protons from a sample also removes informative spin-1/2 nuclei.

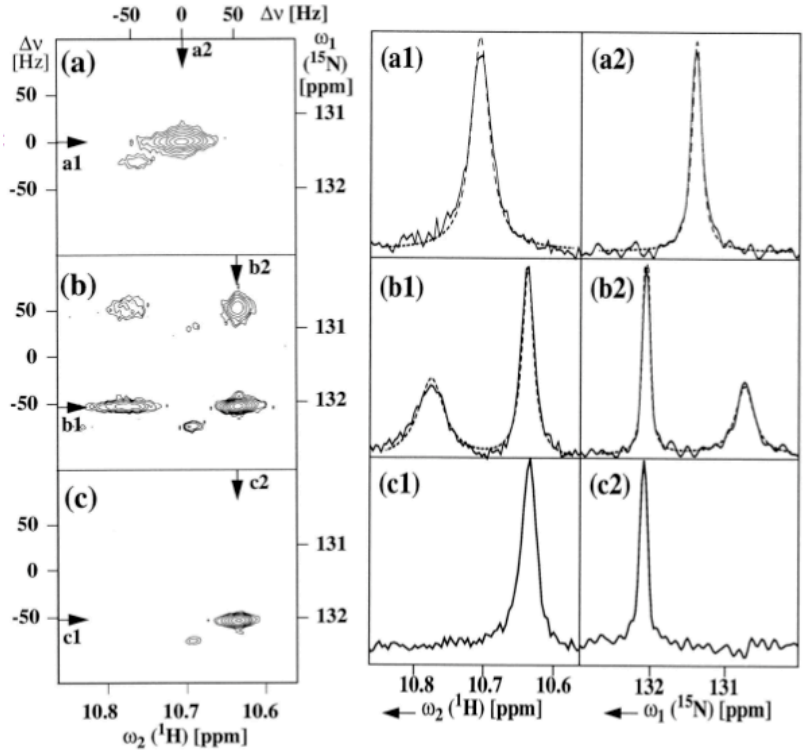
Fortunately, the most-utilized spins in a peptide chain are the amide groups of the protein backbone, which exchange quickly with solvent. Proteins that have been expressed in D_2O media but are purified and resuspended in H_2O solution will be almost fully protonated at the amide residues at any physiological pH. It is recommended to refold proteins with denaturants in H_2O solution to ensure that buried residues exchange fully (Kay, 2004), but in most systems this is straightforward and results in an only slight reduction in yield. When these groups are protonated, a wide array of experiments are possible, including HSQC-like experiments, through-chain experiments such as the HNCA, and through-space amide NOESY experiments.

TROSY has produced a revolution in protein NMR. New versions of NMR pulse programs, including the triple-resonance experiments used for structure determination have been described incorporating the TROSY effect. Some impressively large systems have been studied with TROSY, including the 95% backbone/Cα/Cβ assignment of Malate Synthase G at 81 kDa (Tugarinov, et al., 2000), and the resolution of peaks in 470-kDa and 900 kDa complexes of GroES and GroEL (Fiaux, et al., 2002).

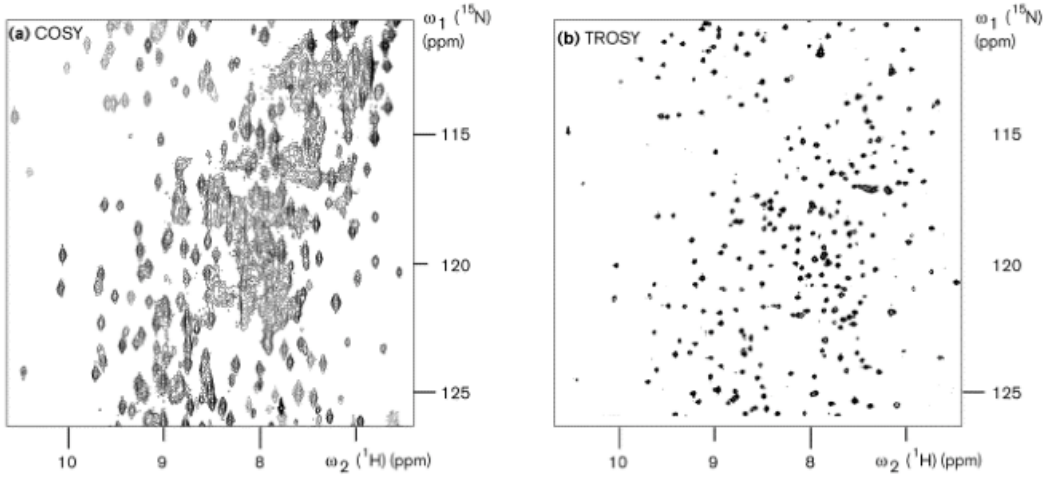
4.2.3.5 Specific labeling strategies

A second disadvantage of large protein NMR, which cannot be overcome with TROSY alone, is the increased incidence of peak overlap. Peak overlap, which prevents the accurate assignment of peaks, has severe effects on the ambiguity of NMR analysis. One technique to address this difficulty is the preparation of specifically-labeled samples, in which specific amino acids or functional groups of interest are isotopically labeled in an NMR-silent background (or vice-versa).

Figure 3. Demonstration of the TROSY effect. a) A decoupled HSQC peak and the respective linewidths in the 1H and 15N dimensions of the fourier-transformed spectrum. b) A non-decoupled HSQC showing the two-dimentional mupltiplet. c) Output data from a TROSY spectrum. From Pervushin, et al., 1997. d) Example of a COSY and TROSY spectrum on a 45-kDa sample, from Wider and Wuthrich, 1999.



d)



The mix-and-match aspect of specific labeling provides a nearly endless array of labeled samples that can be used. However, depending on the information desired, there is a fine balance between the number of labels used and information gained, versus the increase in dipolar relaxation. The simplest case, of $^{1}H/^{15}N$ labels at specific amino acids in a $^{1}H/^{14}N$ background, has been used in small systems, often in combination with fully-labeled samples, to disambiguate assignments (Muchmore, et al., 1989). This can be adjusted to larger systems by utilizing $^{1}H/^{15}N$ labels in a deuterated, $^{2}H/^{14}N$ background (Arrowsmith, et al., 1990). If full-chain backbone assignments are desired, one can include ^{13}C labels, either only in the residues of interest ($^{1}H/^{13}C/^{15}N$ in $^{2}H/^{12}C/^{14}N$ background) (Metzler, et al., 1996) or in the full protein backbone ($^{1}H/^{13}C/^{15}N$ in a $^{2}H/^{13}C/^{15}N$ background). This second labeling scheme would enable full-chain backbone assignment by experiments such as the HNCA, but produce sparser, more easily identifiable peaks in ^{1}H-based TOCSY and NOESY spectra. Unfortunately, the ^{13}C-^{1}H interaction in the sidechains of this labeling scheme is known to cause line-broadening of both the ^{13}C and ^{1}H due to dipolar relaxation, so this labeling scheme is restricted to smaller systems (Kelly, et al., 1999).

A particularly successful technique is specific methyl protonation, which utilizes the protonation of γ2 methyl groups on Ile, Leu, and/or Val sidechains in a deuterated background. This labeling scheme has been used extensively to acquire through-space NOE restraints in hydrophobic cores of proteins, as well as to probe for drug-binding sites and protein-protein interaction surfaces (Goto and Kay, 2000; Kay, 2004).

Specifically labeled proteins are expressed by adding isotopically-labeled amino acids or amino acid precursors to the media prior to induction. However, isotopic labeling is complicated by the tendency of amino acids and precursors to enter unwanted biosynthetic pathways and scramble during cell growth, diluting the labeled from its desired location, and distributing it elsewhere. A number of techniques exist to combat this problem, including the use of high-yield cell-free synthesis systems (Spirin, et al., 1988; Staunton et al., 2006), expression in *E.coli* cells with lesions in certain biosynthetic pathways, or the addition of NMR-silent metabolic precursors that can be used to saturate unwanted biosynthetic pathways (Muchmore, et al., 1989). This scrambling effect is amino-acid dependent. In *E. coli*, those residues that lie at the end of anabolic metabolism pathways are the least likely to scramble. For example, the incidence of scrambling of the backbone ^{15}N atom is low for C, H, K, M, N, P, and R residues; but it is high for E and Q (Muchmore, et al, 1989). Specific labeling protocols must take these sequence-specific variations into consideration.

One labeling protocol of interest for this thesis is the example of backbone assignments being performed on a 47-kDa homodimer with $^{1}H/^{14}N$-labeling of F, Y, T, I, and V residues in a fully-deuterated ^{15}N-labeled background (Kelly, et al., 1999). Since this labeling scheme does not require ^{13}C, it has the dual advantages of avoiding line-broadening due to ^{13}C dipolar coupling and also reducing the expense of the media. The authors accomplish backbone assignments by relying on $^{14}N/^{15}N$ half-filtered NOESY, and ^{1}H-TOCSY experiments.

In my attempts to determine structure in the $\sigma^{N}3$ region, I prepared a number of different isotopically-labeled samples, including uniformly (U)-$^{1}H/^{15}N$, U-$^{2}H/^{13}C/^{15}N$, and specifically-labeled $^{1}H/^{15}N$-IG/IF in $^{1}H/^{14}N$ background, and a sample of a $^{1}H/^{15}N$-Ile-Phe

labeled sample in a $^{2}H/^{14}N$ background. I analyzed these samples by standard HSQC and TROSY-HSQC pulse programs.

4.3 Materials and Methods

4.3.1 Expression and purification of isotopically-labeled protein samples

The σ^{54} 69-398 construct was previously cloned into a pET21b3-2 (Pharmacia Biotech) vector by Dr. Eunmi Hong. The σ^{54} 205-398 construct, cloned into a pskb2-10 vector, was kindly provided by the Darst lab. Five different preparations of uniformly labeled ($^{1}H/^{15}N$; $^{2}H/^{15}N/^{13}C$) and specifically labeled (IF/IG – $^{1}H/^{15}N$ in $^{2}H/^{14}N$ background) were produced with different media additives. These preparations are summarized in Table 1. All preparations were grown by adding 1 mL of an overnight starter culture in LB media to 1 L of M9 media including 13.0 g KH_2PO_4, 10.0 g K_2HPO_4, 9.0 g Na_2HPO_4, 2.4 g K_2SO_4, 1.0 g NH_4Cl (^{15}N for U-^{15}N labeled samples), 2.0 g glucose (^{13}C D-glucose for $^{1}H/^{13}C$ samples, U-^{12}C-U-^{2}H-D-glucose for $^{2}H/^{12}C$ samples), 34 mg thiamine, 952 mg $MgCl_2$, 100 mg Ampicillin, 35 mg chloramphenicol, and the contents of 1 mL of M9 100X metal solutions, lyophilized. Where indicated, 1.0 g of $^{2}H/^{14}N/^{12}C$ Isogro (Sigma-Aldrich) was included in the media. Media was grown to an OD of 0.5-0.6. If specific labels were being used, specifically-labeled amino acids were added (100 mg/L) one half-hour before induction. Cells were grown for 5-7 hours with shaking at 37ºC, pelleted by centrifugation at 5,000 rpm, and resuspended in a buffer containing 50 mM sodium phosphate, 500 mM NaCl, 1 mM TCEP, 10% glycerol, and protease inhibitor cocktail (Roche). Resuspended pellets were frozen at –80ºC until purification. To purify, samples were thawed on ice, sonicated at high power for 6 30-second intervals, and heated to 80ºC in a hot-water bath for 30 minutes to precipitate E. coli proteins. Samples were then ultracentrifuged at 30,000 rpm for 30 minutes, and purified with a Ni-NTA 6x-His purification column (Qiagen). Samples were incubated with the resin with rocking for 5 minutes, then washed with 50 mL of a buffer containing 50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0. Resin was eluted with a buffer containing 50 mM Tris-HCl, 300 mM NaCl, and 400 mM imidazole, pH 8.0. Samples were judged to be > 95% pure by SDS-PAGE gel analysis and used without further purification. Samples were switched into final buffers including 50 mM sodium phosphate pH 6.5, 300 mM NaCl, 1 mM DTT, 1 mM EDTA, and protease inhibitor cocktail. Glycerol was added to some NMR samples in a range of between 0-5%. His tags were not removed for further NMR studies.

Table 2. Approximate composition of ISOGRO.

ISOGRO COMPOSITION:		**Amino acid analysis**			
Salts:	30%	Ala	13.4%	Lys	5.6%
Water:	3%	Arg	3.3%	Met	3.0%
Glucose:	1-3%	Asp	13.7%	Phe	4.3%
A.A.'s/Peptides:	65%	Glu	10.2%	Pro	4.3%
		Gly	11.7%	Ser	4.2%
		His	1.3%	Thr	5.0%
		Ile	4.4%	Tyr	3.2%
		Leu	8.0%	Val	6.0%

Table 1. Preparations of σ^{54} 69-398 constructs with media supplements and final protein yield. AAs indicates supplemented remaining unlabeled amino acids, 100 mg/L of media. All samples were grown in 1 L of the M9 media specified above.*

	uniform isotopic labels	**Additional specific labels**	**media supplements**	**time to induction (hours)**	**final yield (mg/L)**
Prep 1	1H/15N	--	1H-glucose	~6	8
Prep 2	2H/15N/13C	--	2H-glucose	~36	0.2
Prep 3	2H/15N/13C	--	1H-glucose + 1g isogro	~7	5
Prep 4	1H/14N	IF: 15N	1H-glucose + AAs* + 15N-IF	~7	10
Prep 5	1H/14N	IG: 15N	1H-glucose + 15N-IG	~7	2
Prep 6	2H/14N	IF: 1H/15N	2H-glucose + IF –1H/15N + 1 g isogro	~ 7	7

4.3.2 NMR experiments

HSQC experiments using the Fast HSQC (*fhsqc*) pulse program (Mori, et al., 1995) and TROSY-HSQC (Nietlispach, et al., 2005) experiments were performed on samples in a variety of conditions, in order to optimize data quality. This includes sample concentrations ranging from 100 μM to 0.7 mM, pH 6-8.5, glycerol from 0%-5%, and temperatures between 25ºC and 50ºC. HSQC/TROSY experiments were generally run with 1024 and 256 points in the ^{1}H and ^{15}N dimensions, respectively, and between 64 and 256 scans, depending on sample concentration. For conciseness, only the best spectra for each isotopic labeling strategy are discussed below.

4.4 Results

4.4.1 Optimization of production strategies of isotopically labeled samples

One of the difficulties in producing isotopically labeled protein samples can be significantly reduced yields in combination with the high expense of deuterated M9 media. I encountered this problem when trying to produce labeled σ^{54} 69-398 and, therefore, optimized the expression protocol. In two preparations (Prep 2, Prep 3) of U-$^{2}H/^{15}N/^{13}C$ σ^{54} 69-398, I found that adding Isogro, a rich media supplement (Table 2), is key to improving yields (see Table 1). This was also true in making specifically-labeled samples. Although using Isogro in a specifically labeled sample carries the risk of diluting the label of interest, the observed boost in cell growth overcompensated this effect. As suggested by the differential yields and induction patterns of Prep 4 and Prep 5, the supplementation of unlabeled amino acids to M9 media, even in fully protonated media, improves growth. Since it is prohibitively expensive to supplement media with every deuterated amino acid in order to make a specifically-labeled $^{1}H/^{15}N$ protein with a U-$^{2}H/^{14}N$ background, the addition of Isogro is a cost-effective alternative.

4.4.2 NMR on the uniformly 15N labeled σ^{54} 69-398 construct

The HSQC spectra of previously-determined σ^{54} domains $\sigma^{N}2$ and $\sigma^{N}4$ in the DNA-bound and unbound states are shown in figure 5. My objective was to produce high-quality HSQC data of the σ^{54} 69-398 construct, identify peaks that originate from $\sigma^{N}2$ and $\sigma^{N}4$ domains based on their chemical shifts, and assign the remaining peaks from the $\sigma^{N}3$ region.

It should be mentioned that I also studied U-$^{1}H/^{15}N$ 205-398 in both the free and DNA-bound forms, but this construct was less soluble than the 69-398 construct, prone to aggregation, and gave very poor spectra (data not shown). Subsequent work was, therefore, done exclusively on the 69-398 construct.
In the following description of NMR spectra each NMR sample is treated separately.

4.4.2a. NMR on the 69-398 $^{1}H/^{13}C/^{15}N$ protein

To study σ^{54} 69-398 I initially recorded a series of HSQCspectra using a $^{1}H/^{13}C/^{15}N$-labeled sample (figure 6). Although a number of intense peaks are visible, a region with considerable line broadening is present in the middle of the spectrum, and the number of peaks counted is ~ 80 instead of the full 340 expected. Data quality was found to depend on temperature, with the best data being taken at 323 K. However, even at this temperature, considerable line broadening is evident. If this line-broadening is due to unfolded or unstructured regions of the protein, it may be reduced by forming the protein-DNA complex. Therefore, this sample was titrated into double-stranded 21mer DNA, with a 5% excess of DNA. The data quality obtained with the protein-DNA complex sample was extremely poor (figure 7a), most likely due to the large molecular weight (~50 kDa).

4.4.2b. NMR on 69-398 $^{2}H/^{13}C/^{15}N$

To improve data quality, a deuterated sample was made for TROSY spectroscopy. The ^{2}H sample gave much improved data (figure 6b). The protein-DNA complex was also formed and analyzed by TROSY spectroscopy, providing a dramatically improved

spectrum in comparison to the non-deuterated protein (figure 7b). However, the number of peaks in the free (160) and DNA-bound (150) spectra is still only half the number expected (329). To determine which of the peaks in these spectra originate from the σ^N2 and σ^N4 domains, overlays of σ^N2, σ^N4, and 69-398 HSQC/TROSY spectra were performed (Figure 8). While a few peaks overlaid perfectly, many peaks of the σ^N2 and σ^N4 spectra could not be matched up with an equivalent in the spectrum of σ^{54} 69-398. This is likely due to the chemical shift perturbations that originate from the σ^N2 and σ^N4 regions packing against the σ^N3 region. However, nearly every peak in the 69-398 construct falls near a peak in either the σ^N2 or σ^N4 domains, which have a combined 179 peaks.

To overcome the difficulties in transferring chemical shift assignments to the σ^{54} 69-398 construct, TROSY-HNCA experiments (Schulte-Herbrueggen and Sorensen, 2000) were recorded (free and DNA-bound), but signal quality was intractably poor in all conditions tried. I, therefore, applied specific labeling strategies to study the σ^N3 region.

4.4.2c. NMR on 69-398 $^{15}N/^{1}H$ IF/IG in $^{1}H/^{12}C/14N$ background

In order to optimize the expression of specifically-labeled constructs, two fully-protonated samples were made with ^{15}N-labeled isoleucine/phenylalanine and isoleucine/glycine in a $^{1}H/^{14}N$ background. The HSQC spectrum is of poor quality, but shows (figure 10) clusters of peaks in the central and upper left regions of the spectrum, as expected for the average chemical shifts of I, F, and G residues.

4.4.2d. NMR on 69-398 IF in $^{2}H/^{12}C/^{14}N$ background

Based on a refined expression protocol, a $^{1}H/^{15}N$-I/F-labeled sample in a $^{2}H/^{14}N$ background was prepared (Figure 11). In addition to showing considerably reduced spectral complexity, the data quality in this spectrum is considerably improved over the protonated sample used previously (3.3.2c). The total peak count (28) is closer to the expected number (35) than in the case of uniformly labeled samples. However, although data quality was much improved, further optimization would be required to improve the signal-to-noise and linewidths in these spectra.
Peak-counting statistics for all samples described above are summarized in Table 3.

Table 3. Peak-counting statistics from HSQC/HSQC-TROSY spectra.

Sample labeling scheme	**# of backbone amides per domain**			**Total # backbone peaks expected**	**Total # peaks observed**	**% of expected peaks observed**
	σ^N2	***σ^N3***	***σ^N4***			
Prep 1: U-^{1}H/^{15}N/^{13}C	83	150	67	329	~80	~24%
Prep 1b: U-^{1}H/^{15}N/^{13}C + DNA	83	150	67	329	~10	~3%
Prep 3: U-^{2}H/^{15}N/^{13}C	83	150	67	329	~160	~48%
Prep 3b: U-^{2}H/^{15}N/^{13}C + DNA	83	150	67	329	~150	~46%
Prep 6: ^{1}H/^{15}N-IF in ^{2}H/^{14}N background	7	20	8	35	~28	~80%

Figure 5. Previously determined HSQC spectra of a) the core-binding domain (69-198), b) free DNA-binding domain (300-398), and c) DNA-binding domain in complex

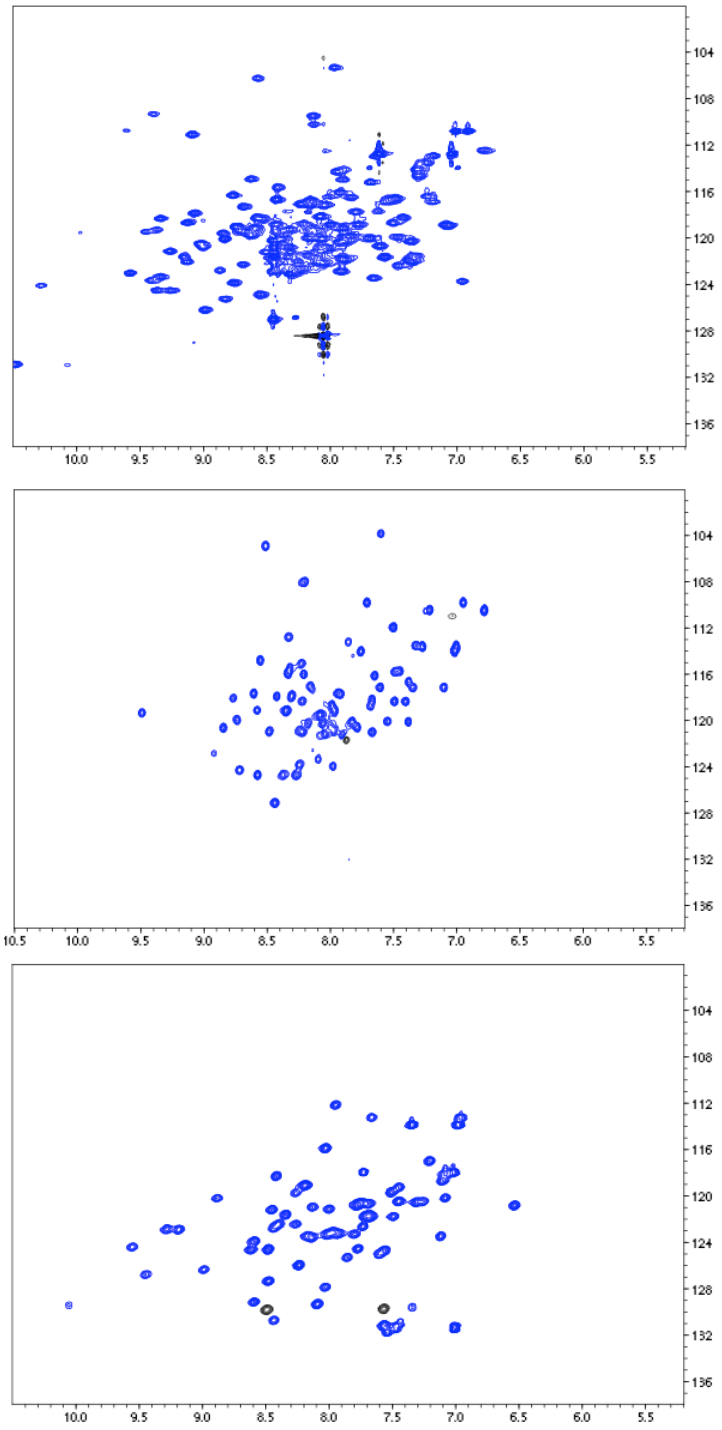


Figure 6. a) ^{15}N-HSQC spectra of the 69-398 construct using ^{1}H/^{15}N labeling and b) ^{15}N-TROSY-HSQC, with ^{2}H/^{15}N labeling. Sample concentration is 0.25 mM, temperature 323 K.

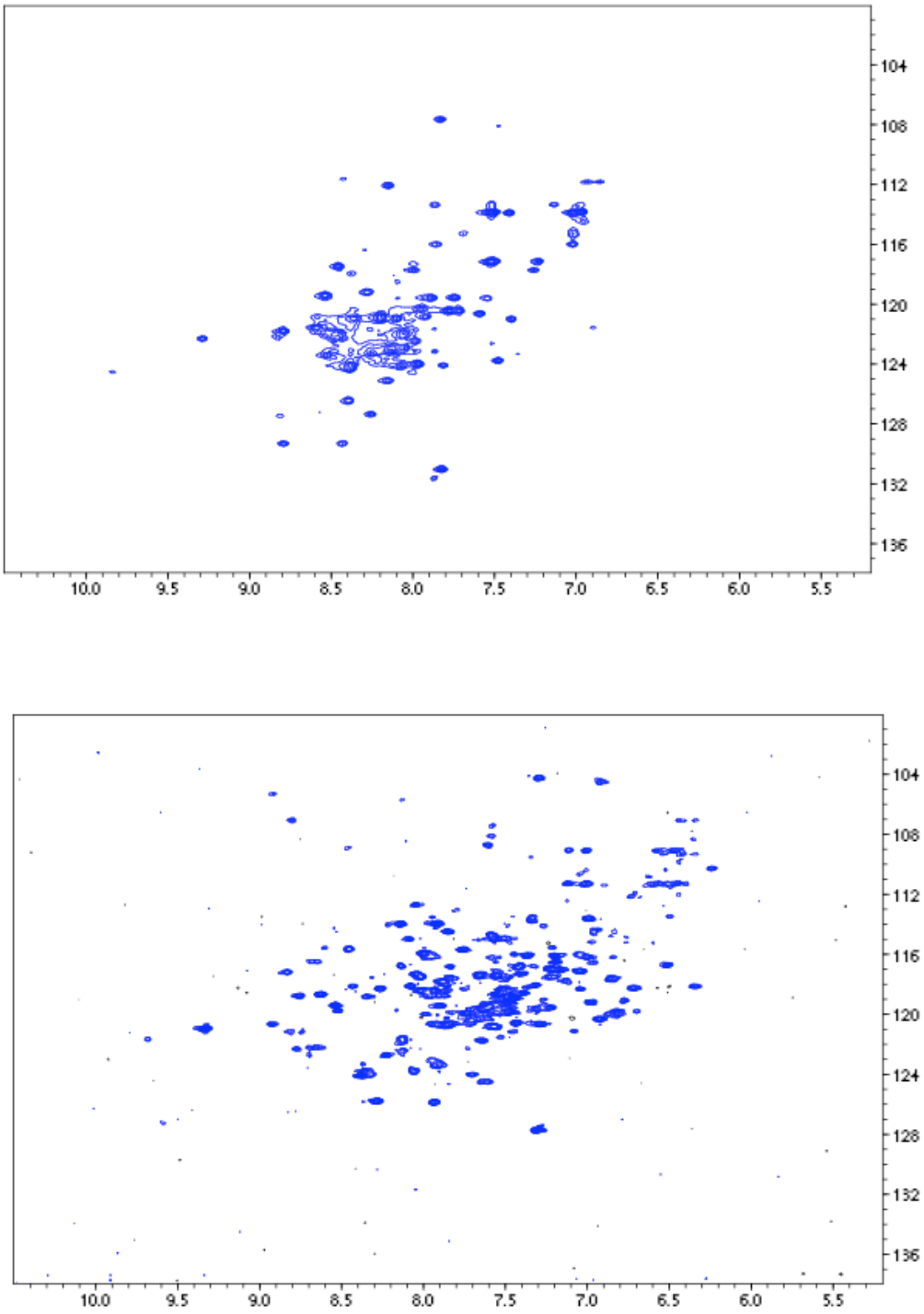


Figure 7. a) 69-398 construct in complex with DNA with 1H/15N labeling b) and with 2H/15N labeling. 10 hr spectra, sample conc. ~ 200 µM.

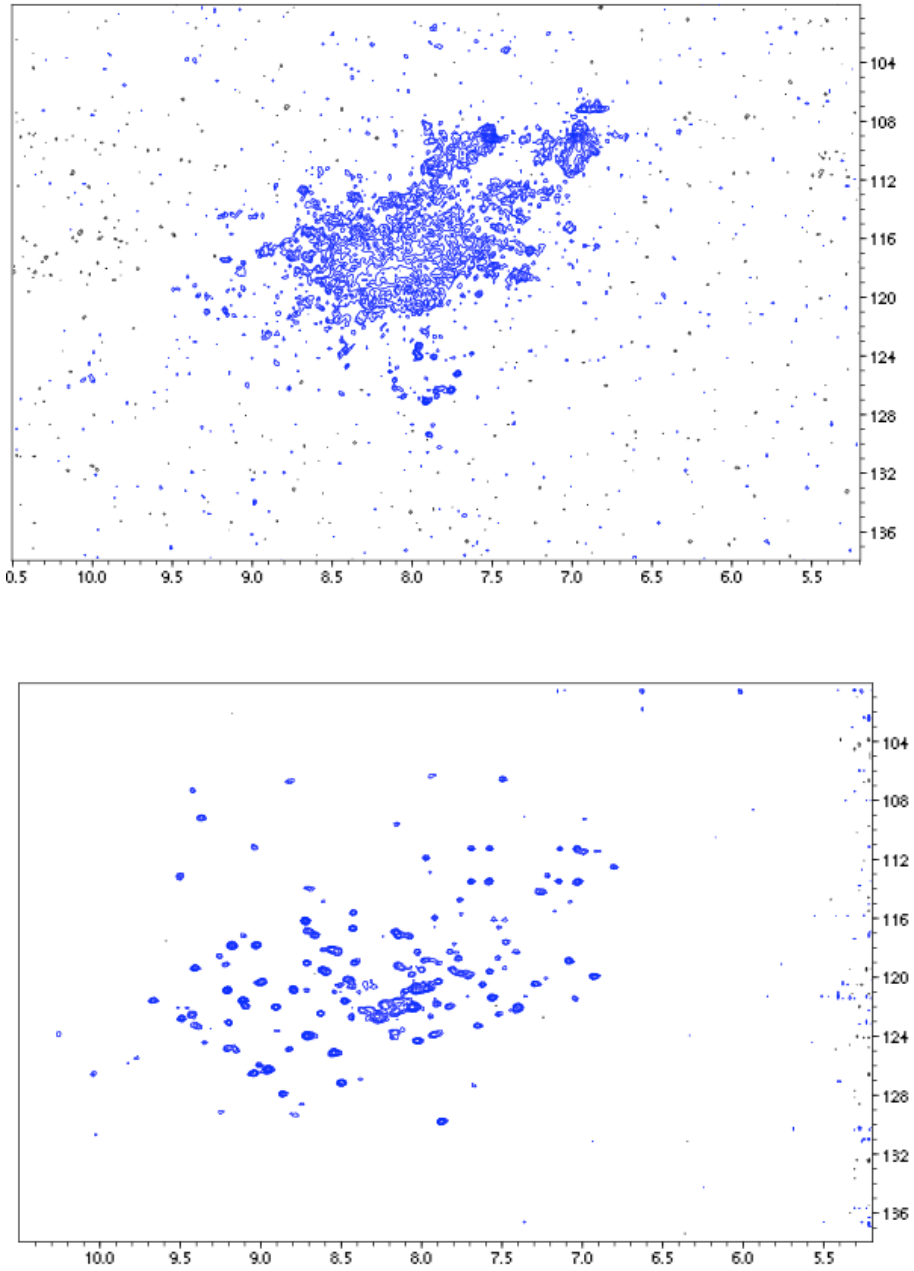


Figure 8. Overlay of U-^{2}H/^{15}N HSQC spectra of the 69-398 construct in the free (blue) and DNA-bound (red) states.

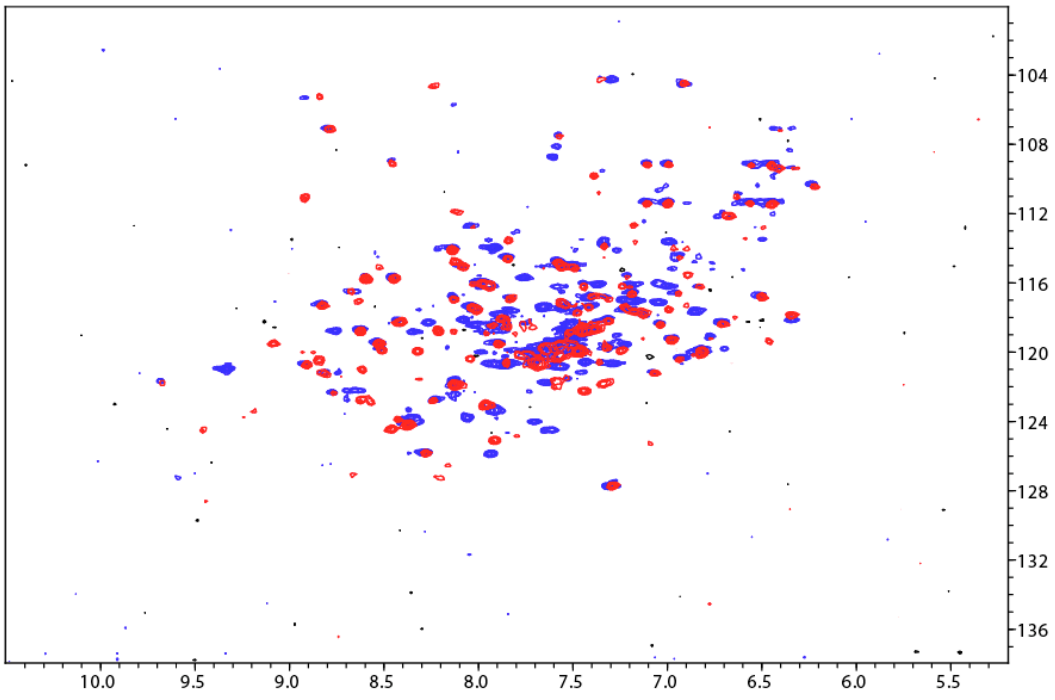


Figure 9. Overlays of ^{15}N-HSQC spectra of σ^{54} 69-398. a) free U-^{2}H/^{15}N 69-398 with core-binding domain (orange), b) free U-^{2}H/^{15}N 69-398 with free DBD (red), c) bound ^{2}H/^{15}N 69-398 with bound DBD (green).

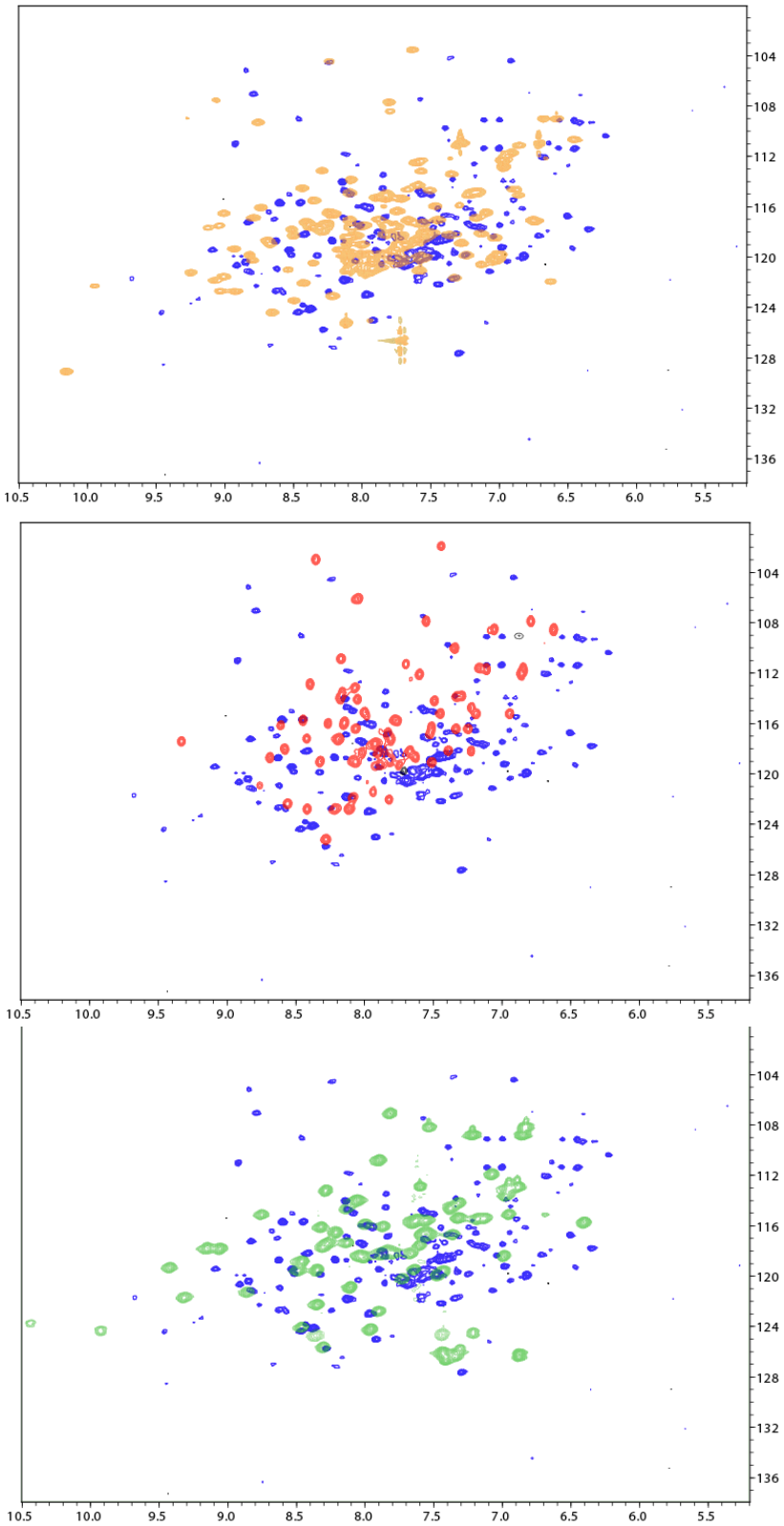


Figure 10. a) 69-398 construct with IF-residues and b) IG-residues $^{1}H/^{15}N$ labeled in a $^{1}H/^{14}N/^{12}C$ background.

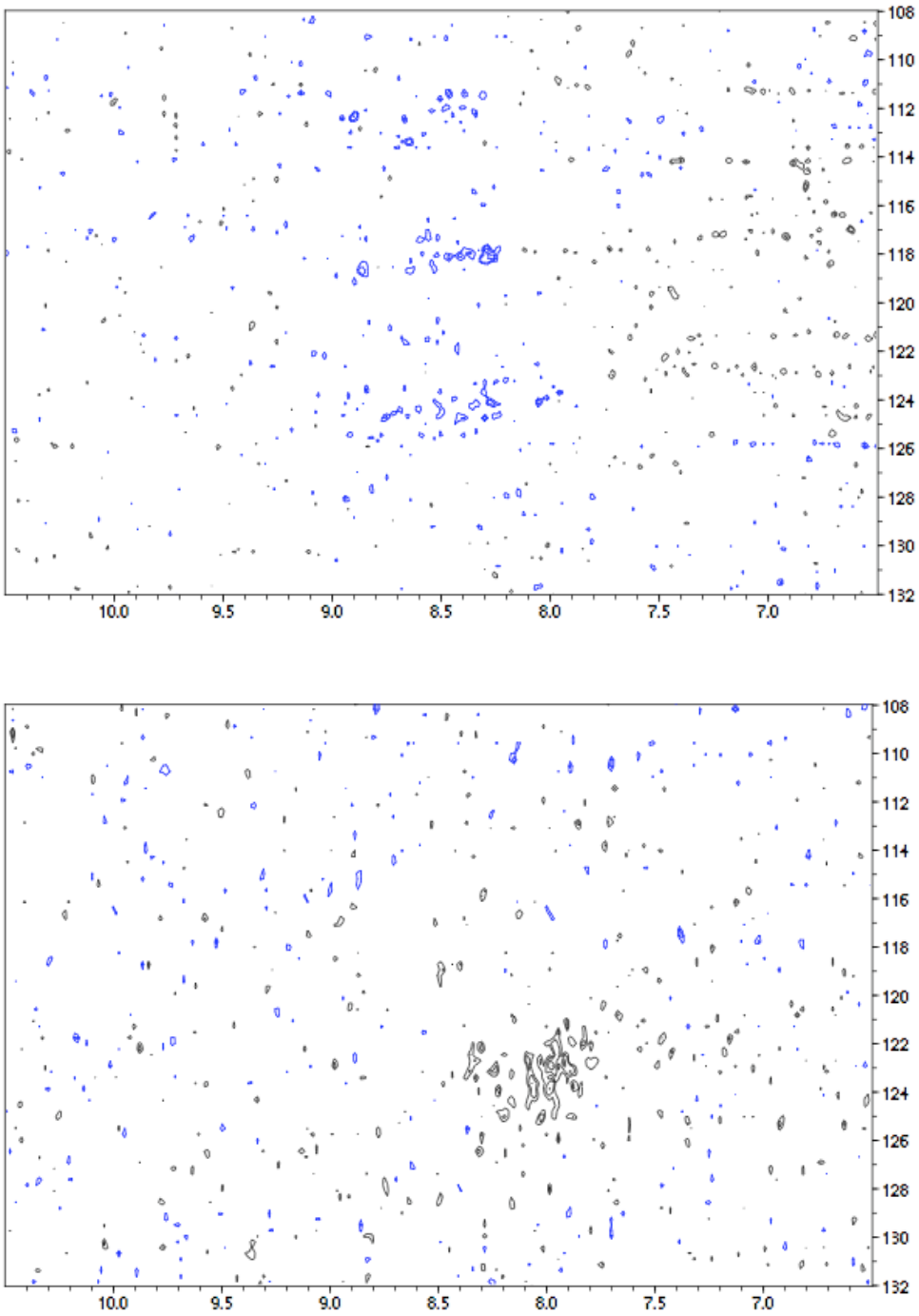
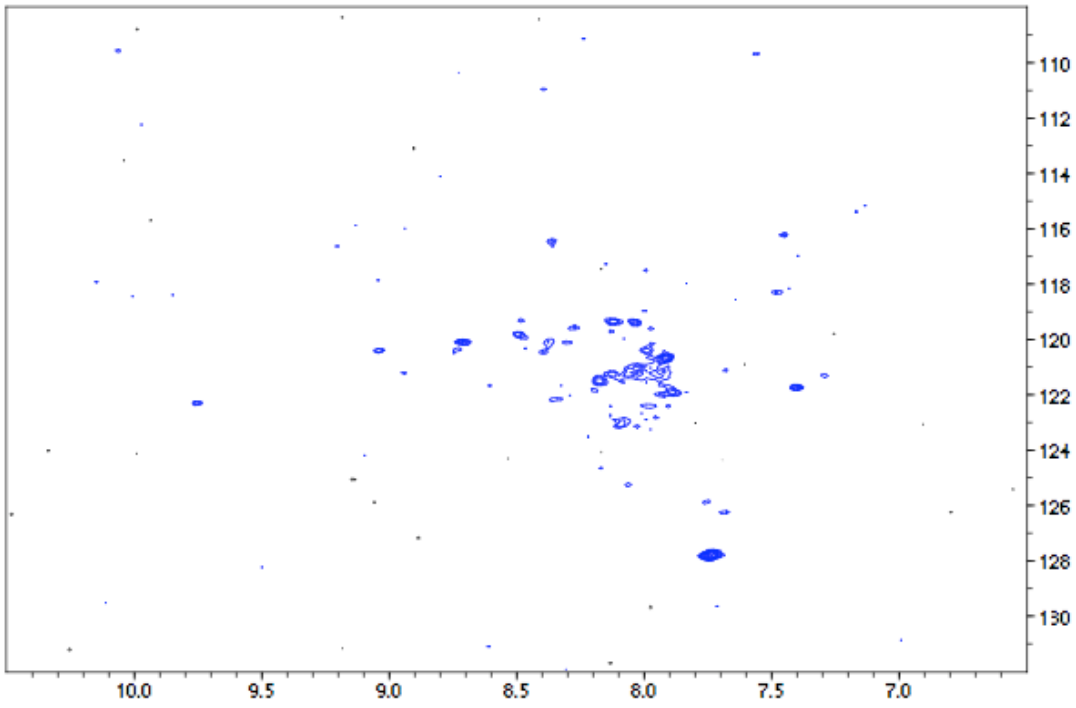


Figure 11. 69-398 construct with IF residues $^{1}H/^{15}N$-labeled in a $^{2}H/^{12}C/^{14}N$ background.



4.5 Discussion

4.5.1 Spectral quality and protein aggregation

The spectra shown in figures 6-11 provide excellent examples of the power of TROSY spectroscopy. However, the quality of the TROSY spectra is still too low to make firm conclusions or collect NOESY/TOCSY datasets on σ^{54} 69-398. The small number of peaks observed in either the uniformly labeled samples or the specifically-labeled samples suggests the presence of protein aggregation. This is consistent with the observation that NMR spectra of all samples worsened over the course of ~1 week, and samples over 200 μM formed precipitates over the course of 1-2 days, regardless of their storage temperature. While short HSQC and TROSY spectra could successfully be recorded, both degradation and aggregation of σ^{54} 69-398 precluded the acquisition of triple-resonance experiments, such as 3D NOESY experiments.

The structural origin for the aggregation of σ^{54} 69-398 may lie in its potentially high surface hydrophobicity. In a functional environment, the σ^N2 and σ^N3 regions of σ^{54} are predicted to interact with RNAP and possibly region σ^N1 (Bose and Zhang, 2008). In the cryo-EM density of σ^{54} in complex with PspF presented in section 1.4 (figure 4), the regions assigned to region σ^N2 and σ^N4 protrude from the RNAP holoenzyme, while region σ^N3 is partially buried in the RNAP density (Chapter 1; Figure 4). Aggregation may also promoted by disorder in σ^N3, as suggested by missing electron density observed in this region by the Darst lab.

The data shown in figures 6b and 7b, however, do show a large number (~150) of sharp, well-resolved peaks. It is possible that these intense peaks come exclusively from the σ^N2 and σ^N4 domains. If these domains tumble independently, or semi-independently of the σ^N3 region, they may experience less line-broadening. This would produce spectra that are completely dominated by the σ^N2 and σ^N4 peaks, without much contribution from the σ^N3 region. However, if this phenomenon were occurring in a pure form; i.e., if the σ^N2 and σ^N4 regions tumbled completely independently of the σ^N3 region, the peaks originating from these terminal domains would be sharp, even in the protonated HSQC samples. The fact that the TROSY effect causes such a notable improvement in data quality suggests that the σ^N2 and σ^N4 regions experience some level of docking against the σ^N3 region.

Based on the structure-in-progress shown in figure 1, it would be reasonable to hypothesize that the DNA-bound form of the σ^{54} activator would experience greater rigidity at the σ^N3 and σ^N4 domains and potentially less aggregation and fewer unstructured regions, due to the -12/-24 interaction with DNA. However, the free and bound forms of the $^2H/^{13}C/^{15}N$-labeled 69-398 samples have similar linewidths and similar numbers of peaks. This suggests that if aggregation or disorder is the causing of the missing peaks, it is present in both free and DNA-bound samples.

The data from the 69-398 $^1H/^{15}N$-IF labeled sample in a $^2H/^{14}N$ background are encouraging, but inconclusive.

4.5.2 Future studies

The main tool to minimize aggregation is sample optimization. Although fairly thorough sample optimization was performed on the σ^{54} $^{1}H/^{15}N$ and $^{2}H/^{15}N/^{13}C$ samples (involving different conditions in over 70 HSQC/TROSY spectra), this was done sequentially with three NMR samples, using the spectral quality as a guide for sample health. Since this exposed the samples to a number of different buffer conditions over the course of a week, it may have sped the process of aggregation, or be influenced by the age of the protein at the time of the buffer change. Although refolding of the $^{2}H/^{15}N/^{13}C$ sample was attempted once with rapid dilution, the refolded sample gave very poor data quality.

The first variable that could be optimized is glycerol concentration in solutions. Whereas the Darst lab uses 10% glycerol in their solutions, I used 0-3% in my solutions. This was based on the NMR rule-of-thumb that glycerol concentrations higher than 3% raise the sample viscosity and produce severe T2 line-broadening. However, consulting the viscosities of glycerol:water mixtures (Appendix A), indicates that sample with 10% glycerol at 30ºC has a viscosity comparable to a 0% glycerol solution at 20ºC. This suggests that higher quantities of glycerol are tolerated for NMR, provided the samples can be analyzed at higher temperatures.

Screening for other stabilizing additives such as detergents might also be rewarding. Suitable solubility screens that could be performed are commercially available and could rapidly be read-out by Dynamic Light Scattering or Gel filtration analysis.

If conditions are found that produce better data quality, a number of different NMR experiments are possible to explore this system further. In the work described above, the deuterated and specifically-labeled samples did not include ^{13}C-labels. Although incorporating ^{13}C-labels into the sample could introduce additional T2 relaxation from C-H and C-C dipolar couplings, they would also dramatically expand the possible NMR experiments. In addition to backbone experiments such as HNCA, HNCO, and HNCACB experiments on a uniformly-labeled ^{13}C sample, specific ^{13}C labels could also be used to hone in on regions of interest. For example, the specific-labeling of IF residues with $^{1}H/^{15}N/^{13}C$ could enable unambiguous assignments of the spin systems in which Ile and/or Phe residues neighbor one another. Ile/Leu/Val or Ala-Methyl-labeling, as described in Section 4.2.3.5, could also give more structural information in hydrophobic regions, provided assignments could be determined.

Chapter 5: Exploring two-component regulatory networks in *Aquifex aeolicus*

5.1 Summary

In this Chapter, I produce a model for correlations between σ^{54} activators and σ^{54}-dependent genes. Since standard techniques usually used to elucidate biochemical pathways are not viable in *A. aeolicus* due to its thermophilicity, we attempt to use the tools at our disposal, including a thorough literature search, gel-shift assays, and consensus-sequence searches, in order to make as much progress as possible. In Part 1 of this chapter, I identify two new activator-gene associations, including NtrC4-*lpxC* and NtrC3-*dhsU*. In the second part of this chapter, I identify a histidine kinase, HksP4, which most likely forms a two-component regulatory system with NtrC3. I find this histidine kinase is a soluble, two-domain protein with a PAS domain that binds heme. Its heme was found to bind O_2, CO, and NO, but not H_2S. However, no forms of the kinase had autophosphorylation activity in ^{32}P-ATP or ATPγS kinase assays, suggesting that activity either requires an additional modulator or protein activator, or physiological conditions.

Part 1: Gene-association studies in *Aquifex aeolicus*

5.1.1 Introduction

5.1.1.1 Introduction to function-correlation studies in *A. aeolicus*

A common complaint of structural biologists is that we spend a great deal of energy determining protein structures without placing the structure in a functional context. Although there is great value in pure structural information, the ability to describe how structural differences in homologues reflect their functional roles often produces greater impact. The previous chapters of this thesis have described structural work on NtrC-like proteins and σ^{54} from *A. aeolicus*. However, the genetic functions of *A. aeolicus* NtrCs were only discussed cursorily, in large part because they are not well understood. There are a number of reasons why a deeper understanding of the functional context of *A. aeolicus* σ^{54} regulation, including the genes regulated by σ^{54} activators and the signals that activate them, would be beneficial.

Unfortunately, the traditional exploration of genetic pathways in *A. aeolicus* is hindered by the organism's extremophilicity. *A. aeolicus*, which was first discovered in sulfurous hot springs such as Yellowstone National Park (Setchell, 1903), grows optimally at 95ºC, making it one of the most extreme thermophiles known. It can grow on hydrogen, oxygen, carbon dioxide and mineral salts, but cannot be grown with sugars,

amino acids, yeast extract or meat extract as a food source (Deckert, *et al.,* 1998). Due to these difficulties, few labs in the world culture *Aquifex aeolicus*. Therefore, most techniques traditionally used to elucidate genetic pathways in prokaryotes, such as the growth of auxotrophic strains and whole-genome microarrays, are not viable. Even basic biochemical assays adjusted to high temperatures are complicated by the use of temperatures that exceed the melting temperature of DNA.

In lieu of traditional assays to trace biochemical pathways in *A. aeolicus,* I have attempted to further our understanding as much as possible with the limited tools available. This includes sequence analysis, gel-shift assays, and an extensive literature search to attempt to find and catalogue possible functional correlations.

5.1.1.2 Determination of gene associations and putative binding sites

My first task was the determination of σ^{54}-promoter regions in the *A. aeolicus* genome. σ^{54}-promoter regions often contain three different characteristics: a σ^{54} conserved binding site centered at –12 and –24 base pairs upstream of the transcription initiation site; two transcriptional activator sites located ~100 bp upstream of the transcription initiation site; and in some cases, an IHF binding site located between the σ^{54} and activator binding sites (Santero, et al., 1992).

Although all three features can assist in the identification of a σ^{54}-dependent gene, the identification of σ^{54}-binding-sites is the most straightforward and prescriptive. As described in Chapter 1, σ^{54} binds to sites with the consensus sequence mrNrYTGGCACG-N4-TTGCWNNw. The mrNrYTGGCACG region is called "-24" region, and tTGCWNNw is called the "-12" region, since they are often located –12 and –24 base pairs upstream from the transcription initiation site (Barrios, et al., 1999). Six σ^{54}-binding sites have been previously identified upstream of the glnB, fhp, nirB, dhsU, aq_1119, and aq_087 genes (Barrios, et al., 1999).

Unlike the strong consensus available for σ^{54} binding sites, DNA-binding sites of activator proteins are poorly conserved and exhibit considerable variation from one activator protein to another. Even consensus sequences specific to one activator, such as the weak consensus sequence for NifA, tend to be very general, such as TGT-N_{10}-ACA (Ray et al., 2002). However, many activator binding sites exhibit certain patterns. Since the DNA-binding domains of σ^{54} activators are symmetrical dimers, palindromic binding sites or sites with pseudo-symmetry are common. A-tracts interrupted by YR base-pair steps, which constitute one of the most deformable motifs known in protein-DNA interactions (Lankas, et al., 2003), are also common, as discussed in chapter 2. A selection of activator binding sites from NtrC, NifA, and *A. aeolicus* NtrC proteins demonstrate both the common features of these binding sites, as well as their diversity (Table 1).

A. aeolicus also contains 5 σ^{54} -transcriptional activators with intact regulatory, central, and DNA-binding domains: NtrC1, NtrC2, NtrC3, NtrC4, and NifA (Deckert, et al., 1998). To attempt to match these activators to σ^{54} UASs and learn about the function governed by σ^{54} in *A. aeolicus*, a combination of consensus sequence searches, gel shift assays, and an exploration of the functions of these putative σ^{54}-dependent genes was performed.

Table 5.1.1. Examples of activator binding sites.

Binding site	Sequence	Reference
NtrC (E. coli) high affinity	ttgcaccaacatggtgctt	
NtrC (E. coli) low affinity	aagcactatattggtgcaa	(Atkinson, et al., 2002)
NifA (K. pneumoniae)	**aca**ccctttgtct**tgt**	
	acagcggacaagt**tgt**	
	acagaataaggtg**cgt**	(Ray et al., 2002)
NtrC1 (*A. aeolicus*) high affinity	aattgccaaaatggcaa	
NtrC1 (*A. aeolicus*) low affinity	ttttgtcagactggcaaa	(Doucleff, et al., 2005)
NtrC4 (*A. aeolicus*) high affinity	ctttttgcaaatttgcaaatgcat	
NtrC4 (*A. aeolicus*) low affinity	gcgattggtaaaactacaaaagag	(Thesis Chapter 2.)

5.1.2. Materials and Methods

5.1.2.1 Protein expression and purification

Plasmids containing full-length constructs of NtrC2 and NtrC4 cloned into Pet21a plasmids were provided by Joe Batchelor. The full-length protein of NtrC3 was cloned into a Pet21a vector. The NtrC3 gene was amplified by PCR from genomic DNA in *Aquifex aeolicus*, cut with NdeI and BamHI restriction enzymes, and ligated into a PSKB3 plasmid using DNA T4 Ligase (NEBiolabs). All vectors contained an N-terminal His_6 tag and carried kanamycin resistance.

Plasmids were transformed with a TFB-chemically competent transformation into *E. coli* cells with a Rosetta.pLysS plasmid with chloramphenicol resistance. All proteins were grown by the auto-inducing Studier growth method (Studier, 2005). 2x1 L cultures of Studier media with supplemented 50 mg/L kanamycin and 35 mg/L chloramphenicol were inoculated with 10 mL overnight starter culture Luria Broth media. Flasks were shaken at 37ºC for 24-48 hours, then pelleted at 5,000 rpm for 30 minutes. Resulting pellets were resuspended in a buffer containing 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM TCEP, 10% glycerol, and one tablet of Roche protease inhibitor cocktail. Samples were sonicated at 70V in conical flasks at 4ºC for 6x30s intervals, then heated in a water bath to 70ºC for 30 minutes. Precipitated *E. coli* proteins and insoluble cell matter were removed by ultracentrifugation at 30,000 rpm for 30 minutes. The supernatant was filtered to 0.4 µm with VWR sterile filters, and flowed over a Qiagen Ni-NTA column. The column was washed by a wash buffer containing 50 mM Tris-HCl, 300 mM NaCl, and 20 mM imidazole, and eluted with the same buffer containing 400 mM imidazole. 50 mL flow-through, wash, and elution fractions were collected for subsequent SDS-PAGE gel analysis. TCEP, 10% glycerol, and protease inhibitor were immediately added to the eluent fraction. Eluent was dialyzed overnight into the final storage buffer, 50 mM

Tris-HCl, 300 mM NaCl, 1 mM DTT, 1 mM EDTA, and 10% glycerol with 12,000 molecular weight cutoff dialysis membrane. The 6x-histidine tag was not removed for subsequent DNA-binding applications. Samples were concentrated to 5 mg/mL for DNA binding assays.

5.1.2.2. Consensus sequence searches

Protein domain architecture was predicted with a combination of p-BLAST (amino acid sequence) and n-BLAST (nucleotide BLAST). Consensus sequence searches through the *A. aeolicus* genome were performed with PATSER (Dombrecht, et al., 2002).

5.1.2.3. PCR, cloning of binding sites into Pet21a vector

The 300-base pair sequences upstream of the fhp, dhsU, nirB, glnB, lpxC, and Aq_087 sites were amplified by PCR using primers designed by Johanna Heideker. Since PCR with a genomic template is not as efficient as PCR with a plasmid template, UAS regions for these six genes were cloned into Pet21a vectors. The PCR inserts listed above were additionally amplified with NdeI and BamHI cut sites on each end. Sites were cut with NdeI and BamHI enzymes for 2 hours at 37ºC, Qiagen Mini-prep purified, and ligated into a cut Pet21a vector with T4 ligase (New England Biolabs) for 16 hours at 6 ºC. Vectors were transformed into Top10 highly competent cells (Invitrogen) and plated on agar plates containing 100 mg/L ampicillin. Cells from single colonies were grown in LB at 37ºC overnight and plasmid was recovered with Qiagen mini-prep plasmid purification columns. Plasmid sequences were confirmed by sequencing at the UC Berkeley DNA sequencing facility.

5.1.2.4 Native protein-DNA gel-shift assays

The large size of 250-bp DNA regions (estimated mass ~ 140,000) compared to dimeric protein (~100 kDa) complicated the effective visualization of bands in PAGE or standard 1% agarose gels. A number of alternative polyacrylamide acrylamide:bisacrylamide solutions were tested, but gel bands were poorly resolved. 2-3% agarose gels allowed the visualization of DNA-bands, but were susceptible to smearing. The most effective gel-shift assay was found with 1% Low Melting Point Agarose (Promega) gels in Tris, Boric Acid, EDTA (TBE) 1X buffer. Gels were run at 80 V for 2 hours at either room temperature and 4ºC. Gels were stained with SYBR-gold nucleic acid stain (Invitrogen) solution for 30 minutes, and visualized with a UV gel-imagining box.

4.1.3 Results and Discussion

5.1.3.1 Search for genes under σ^{54} control in *A. aeolicus*

Table 5.1.2. Functions of putative σ^{54}-dependent genes in A. aeolicus.

gene	function	reference
dhsU	**Sulfur metabolism:** encodes flavocytochrome C sulfide dehydrogenase (FCSD).	Kostanjevecki, 2000
fhp	**NO detoxification:** encodes for the globin portion of flavohemoprotein,	
glnB	**Nitrogen regulation:** encodes for protein PII, a multi-function cytosolic activator	Magasanik, 1993
lpxC	**Cell wall synthesis**: synthesizes the LpxC protein, necessary for the synthesis of lipid A	Raetz, 2002
nirB	**Nitrite reduction:** encodes for the nitrite reductase NAD(P)H (a.k.a. oxidoreductase) large subunit	Morena-Vivian, 2010

Table 5.1.3. Order of genes in the A. aeolicus genome surrounding σ^{54} putative binding sites and the RPON gene. Red genes have putative σ^{54}-binding sites in the UAS. Blue genes are discussed in the text.

glnB (aq_109)	*nirB (aq_206); fhp (aq_211)*	*dhsU (aq_232)*
Fdx4 (ferredoxin) hfq (host factor I) glnB (nitrogen regulatory PII protein) glnA (glutamine synthetase) amtB (ammonium transporter)	AtpG2 (ATP synthase F1 gamma subunit) aq_204 (hypoth. prot) **nirB (nitrite reductase NAD(P)H large subunit)** cobA (uroporphyric-III c-methyltransferase) trpD2(phosphoribosyl-anthranilate transferase) **fhp (flavohemoprotein)** cynS (cyanate hydratase) glnBi (PII-like protein GlnBi) nasA (nitrate transporter) narB (nitrate reductase narB) nifA	nifA (continued from nirB/fhp) dut (dUTP-nucleotidohydrolase) phpA(polynucleotide phosphorylase/polyadenylase) rpsO (primary rRNA binding protein) aldH2 (aldehyde dehydrogenase) ntrC3 (NtrC protein) hksP4 (histidine kinase) **dhsU (flavocytochrome C sulfide dehydrogenase** soxF (Rieske-I iron sulfur protein) fccB (sulfide dehydrogenase, flavoprotein subunit)
Aq_1119 aq_1113 hksP1 (histidine kinase) ntrc1 (NtrC protein) aq_1119 aq_1120		
Aq_087	*lpxC (aq_1772)*	*rpon (aq_599)*
kdsA (2-dehydro-3-deoyphosphooctonate aldolase) modC (molybdenum ABC transporter permease) **aq_087 (hypothetical prot)** aq_088 (hypoth. Prot) aq_090 (hypoth. Prot)	leuS (leucyl-tRNA synthetase beta subunit) hemD (uroporphyrinogen-III synthase) *lpxC (UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase)* aq_1773 (hypoth. protein) rpmA (50S ribosomal protein L27) glnE (glutamate ammonia ligase adenylyl-transferase)	*PurB (adenylosuccinate lyase)* *Aq_598 (putative protein)* ***RPON (σ^{54} protein)*** *Aq_600 (hypoth. protein)* *Aq_601 (hypoth. protein)* *Aq_603 (hypoth. protein)* *LpxA (acyl-[acyl-carrier-protein]-UDP-N acetylglucosamine acyltransferase)* *glmU (UDP-N-acetylglucosamine pyrophosphorylase*

To confirm and update the selection of σ^{54} binding sites presented by Barrios, et al., the σ^{54}-binding consensus sequence was re-screened in the *A. aeolicus* genome with the program PATSER (Dombrecht, et al., 2002; Hertz and Stormo, 1999). Consensus matches were observed upstream of the glnB, nirB, dhsU, Aq_087 and fhp genes, as published previously (Barrios, 1999). However, the consensus sequence also matched a region upstream of the lpxC gene, which we included in our analysis.

5.1.3.2. Functional roles of *A. aeolicus* σ^{54}-regulated genes

To provide a functional context for these putative σ^{54}-dependent genes in *A. aeolicus*, the functions of each gene product was researched in the literature. The genes encoded by open reading frames Aq_1119 and Aq_087 have unknown functions and are not extensively discussed. A summary of known functions is listed in Table 5.1.2.

In most prokaryotes, it is common for genes involved in similar functions, such as nitrogen regulation or flagellar biosynthesis, to be grouped together into operons. However, the *A. aeolicus* genome exhibits a high level of genetic dispersion. Although some genes, such as those encoding flagellar proteins and ribosomal subunits, are grouped into operons, a number of gene classes that are commonly grouped into operons in other organisms (i.e., tryptophan and histidine biosynthesis pathways) are completely dispersed throughout the genome. In many cases, separate subunits of the same enzyme are coded in different parts of the genome, such as the *gltB* and *gltD* subunits of glutamate synthase. It has been suggested that this tendency towards dispersion may be a result of thermophily (Deckert, *et al.,* 1998).

To determine whether any putative σ^{54}-dependent genes are organized into operons, the sequence of genes in coding regions surrounding the putative sites are listed in Table 5.1.3. Although these descriptions of genome clustering are merely suggestive and do not account for the complex details of operon architecture, a number of interesting features are apparent.

5.1.3.3 Gene-by-gene analysis

5.1.3.3a. NirB/fhp

Nitrate metabolism, the breakdown of NO_3- into NH_4^+, NO, N_2O, and N_2, is a process usually performed by multiple unrelated pathways in an organism, which can be grouped into having two distinct functions: nitrate assimilation and nitrate dissimilation. *Nitrate assimilation* is the process that converts nitrate to ammonium for the biosynthesis of nitrogen-containing biomolecules, such as amino acids and cofactors. *Nitrate dissimilation* can be further divided into the categories of *nitrogen respiration,* (anaerobic respiration that uses NO_3^- instead of O_2 as the final electron acceptor in metabolic electron transport chains) and *nitrate reduction* (removal of excess reducing power for redox balancing). All pathways require two or more chemical steps, including nitrate reduction and nitrite reduction (Moreno-Vivian, et al., 1999). A schematic of these reactions is shown in Figure 5.1.1.

According to Professor Valley Stewart of UC Davis (Appendix B), the genes for nirB and fhp appear to be clustered in an operon involved in nitrogen assimilation. The

nirB gene encodes an important gene in nitrite reduction: the nitrite reductase NAD(P)H (oxidoreductase) large subunit. Since NO is a byproduct of nitrite reduction, it is not uncommon for bacteria to include NO-metabolizing enzymes, such as the flavohemoprotein encoded by the *fhp* gene, in assimilative nitrate reduction operons. The cobA enzyme that immediately follows nirB is an enzyme that forms a major branchpoint for heme synthesis that is responsible for the manufacture of siroheme, a cofactor required by the nirB enzyme. The relevance of the trpD2 gene is less clear: although its purported function (synthesis of a tryptophan precursor) is not related to nitrate metabolism, the amino-terminal domain of the TrpD polypeptide comprises the glutamine amidotransferase step of anthranilate synthesis, which converts glutamine to glutamate, and therefore may be related to nitrogen homeostasis (Jackson and Yanofsky, 1974).

The genes following *fhp* also appear to participate in nitrogen assimilation. This includes a *glnB*-like signaling protein, *glnBii*, which acts as a sensor of cellular nitrogen (Magasanik, 1993), nitrate transporters nasA and narB, which are implicated in nitrogen assimilation pathways (Figure 5.2.3; Morena-Vivian, 1999), and the nifA protein, a σ^{54}-dependent activator protein. As described later in this chapter, nifA is expected to activate the *glnB* gene (section 5.1.3.5).

5.1.3.3b. DhsU

The genes for NtrC3 and a histidine kinase, HksP4, are encoded immediately upstream of the dhsU gene (Table 2). This strongly suggests that HksP4 and NtrC3 form a SK-RR pair for two-component signal transduction to activate the dhsU gene.

The dhsU gene, which encodes a flavocytochrome C sulfide dehydrogenase (FCSD), is implicated in the oxidation of sulfur compounds. The known and hypothesized sulfur metabolism pathways from *Chlorobaculum tepidum*, a green sulfur bacterium, is presented in Figure 5.1.2 to provide context for the processes that oxidize/reduce sulfur compounds to fuel photosynthesis. Although some of the same genes are found in *A. aeolicus*, the purpose of sulfur oxidation is chemical, rather than photosynthetic.

The prototypical FCSD catalyzes sulfide oxidation to elemental sulfur, and transfers electrons to cytochrome c (Marcia, et al., 2010). In the *A. aeolicus* genome, it is annotated next to a Rieske iron-sulfur protein called soxF, which is implicated in electron transport (Sakurai, et al., 2010). FCSDs are found in a number of other bacteria, including *Paracoccus denitrificans*, *Thiobacillus sp. W5*, and 8 species of phototrophic green and purple bacteria of the Chlorobia and Chromatiales orders (Kostanjevecki, et al., 2000). Organisms that express FCSD proteins are typically able to utilize reduced sulfur compounds as electron donors for carbon dioxide fixation. In green and purple sulfur bacteria, these electron donors include hydrogen sulfide, as well as elemental sulfur, thiosulfate, and in some cases, sulfite and/or tetrathionate (Kostanjevicki, 2000). A number of chemolithoautotrophic "vent" organisms such as the *Pyrodictium* and *Pyrolobus* genera use hydrogen sulfide for the reduction of O_2 in microaerophilic environments (Stetter, 2006). A recent paper suggests that *A. aeolicus* respiration can utilize the electron donor/acceptor pairs H_2/S^0, H_2S/O_2, or H_2/O_2, via different respiratory pathways (Guiral, et al., 2009). However, since all these pathways are predicted to

reduce sulfur, it is not clear that the FCSD protein would be involved in respiration, or what role it would play if it were. We have also not explored the possibility of the FCSD's involvement in a pathway oxidizing H_2 and reducing sulfate; another observed pathway in extremophiles.

Although the *dhsU* gene appears to be the only sulfur oxidation gene under σ^{54} control, *A. aeolicus* has a number of other genes implicated in sulfur metabolism, including a thiosulfate utilization operon, as well as a sulfide-quinone reductase (SQR), a FCSD homologue (Deckert, et al., 1998; Shahak, et al., 1992; Arieli, et al., 1994). SQR is a membrane-associated sulfide-detoxifying enzyme that, like FCSD, oxidizes sulfide, but it transfers electrons to the membrane quinone pool instead of to cytochrome c (Marcia, 2010).

5.1.3.3c. GlnB

The *glnB* gene encodes a protein called PII, one of the most widely distributed signaling proteins in nature. PII is uridylated by Uridyl Transferase (UTase), which is a direct sensor for alpha-ketoglutarate and adenylylate energy charge that transduces a signal to a number of different receptors, including the NtrB-NtrC TCS in *Klebsiella pneumoniae* and *E. coli* (Jiang and Ninfa, 2009). In *E. coli*, PII becomes uridylated by UTase under nitrogen starvation conditions (low NH_4^+), and loses the ability to interact with and inhibit NtrB, a histidine kinase. This enables NtrB to autophosphorylate and transfer the phosphoryl group to NtrC, which activates transcription at the glnALG operon in a σ^{54}-dependent manner (Hirschman, et al., 1985; Magasanik, 1993). In nitrogen-rich conditions, PII is deuridylated, and interacts with NtrB to inhibit its activity.

In *A. aeolicus*, *glnB* appears to be part of a typical nitrogen regulation operon. The *glnB* gene is followed by the gene encoding glutamine synthetase, *glnA*, an enzyme that catalyzes the condensation of glutamate and ammonia to form glutamine. This is followed by an *amtB* gene, which codes for an ammonium transporter. In a number of organisms, including *E. coli* and *Klebsiella pneumoniae*, the *glnA* gene is under NtrC control (Hirschman, et al., 1985).

The gene preceeding *glnB*, *hfq*, encodes for the Host Factor I protein, an RNA chaperone implicated in the activity of small, regulatory RNAs (sRNA), also known as noncoding RNAs. sRNAs repress translation by base-pairing with the coding or noncoding regions of specific transcribed mRNAs, often by preventing the binding of ribosomes or targeting the mRNA for degredation. Hfq is a chaperone that assists sRNAs to stably base-pair with their mRNA target (Frohlich and Vogel, 2009).

The *hfq* gene is transcribed in the negative direction, and *glnB* is transcribed in the positive direction, meaning that *hfq* and *glnB* share a UAS region. Although *hfq* is not preceded by a σ^{54} activation sequence, its proximity to the *glnB* activation region could suggest a link between σ^{54} activation and an sRNA pathway. There are known examples, including *Vibrio cholerae* and *Vibrio herveyi*, in which σ^{54} has been implicated in controlling expression of sRNAs (Lenz et al., 2004). In future studies of σ^{54}-activation in *A. aeolicus*, especially in the analysis of the UAS of *glnB*, the possibility of a link to an sRNA pathway should be considered.

Figure 5.1.1. Summary of reactions involved in nitrate reduction.

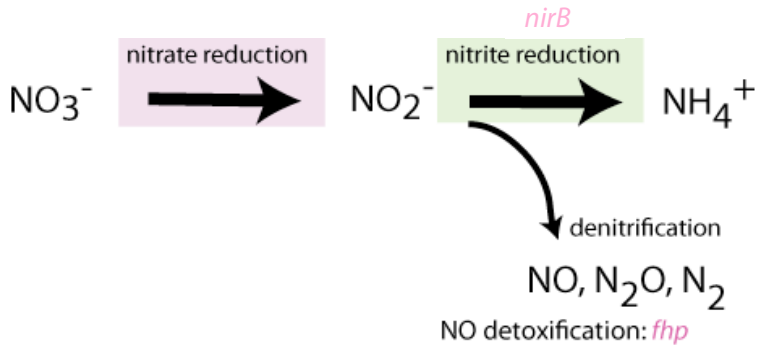


Figure 5.1.2. Overview of known and hypothesized sulfur metabolism pathways of Chlorobaculum tepidum, a photosynthetic green sulfur bacterium. A. aeolicus contains FCSD, SQR, and SoxF genes, highlighted with a red square. From Sakurai, et al., 2010.

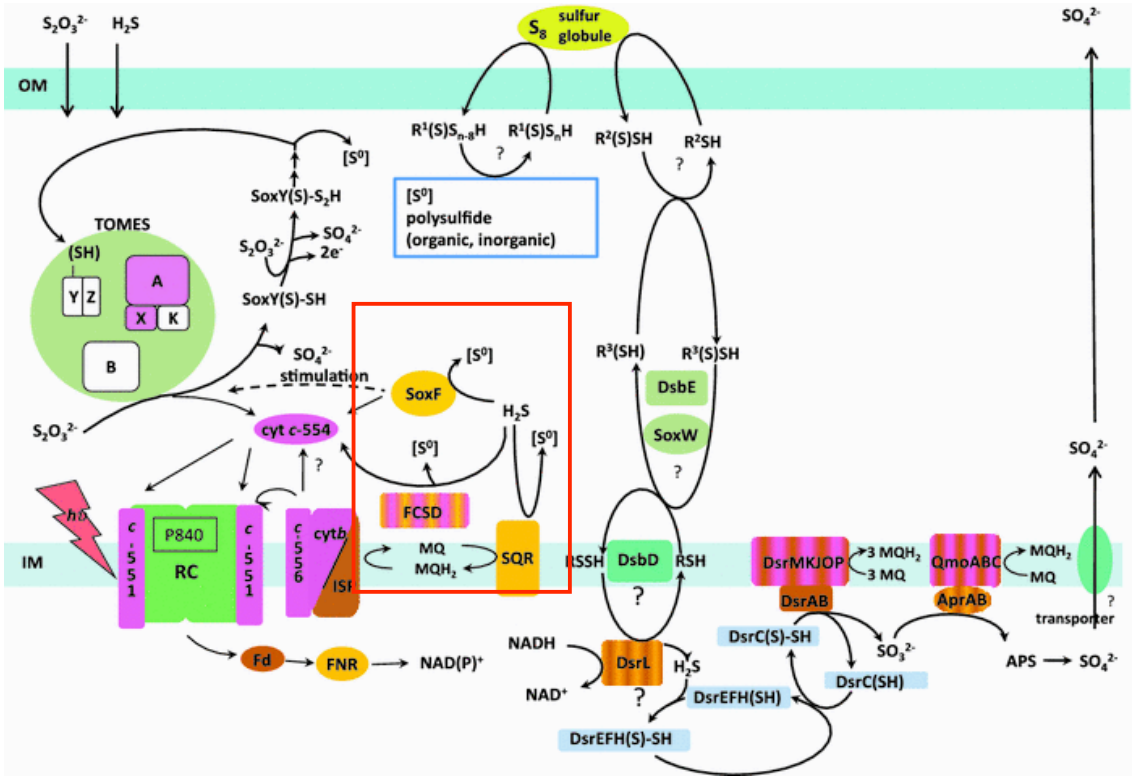
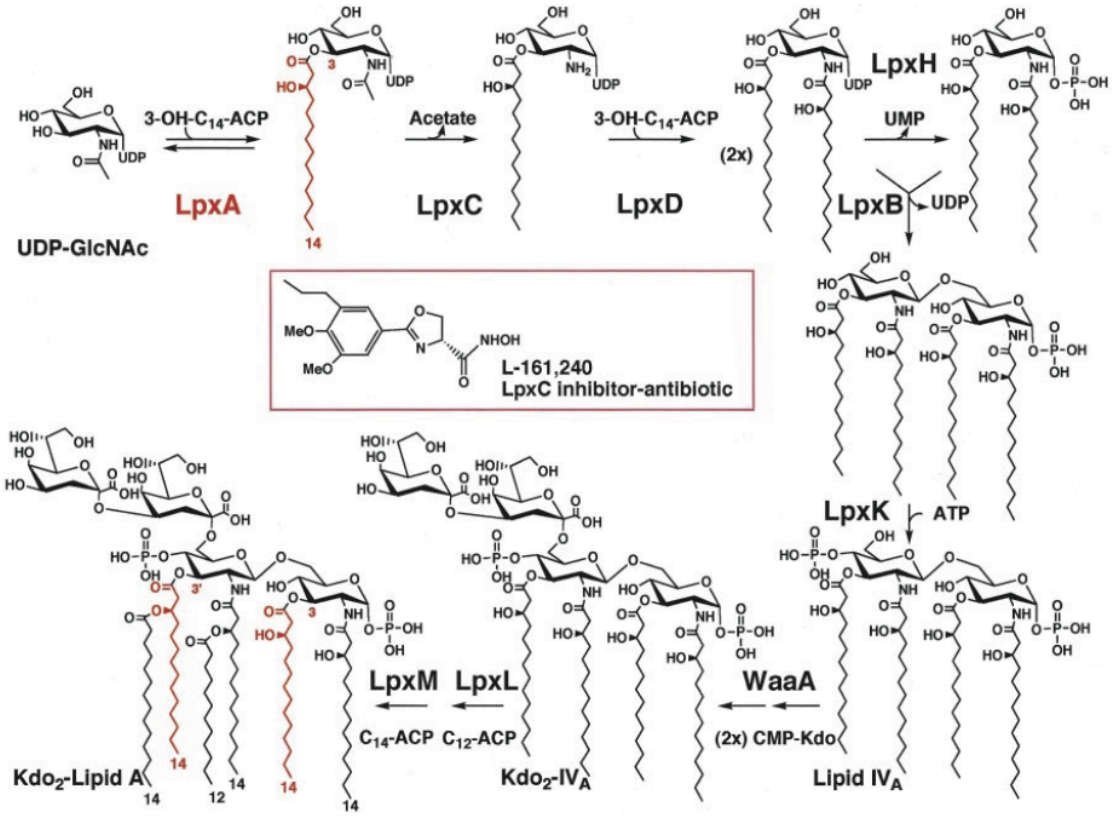


Figure 5.1.3. Biosynthesis pathway of Lipid A. From Raetz and Whitfield, 2002.



5.1.3.3d. LpxC, RPON

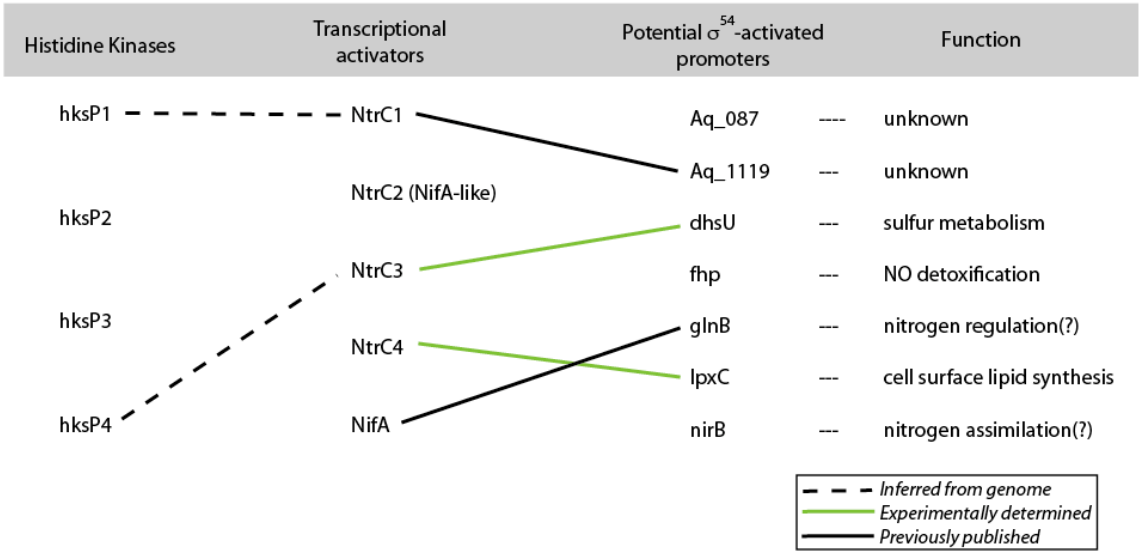
The *lpxC* gene does not appear to be located in an operon. The genes surrounding it, leuS, rpmA, hemD appear to have no direct relation to lipoprotein synthesis. The other genes necessary for the initial steps of the lipoprotein synthesis pathway, lpxA, lpxB and lpxD are scattered elsewhere throughout the genome. Interestingly, *A. aeolicus* does not appear to code for the enzymes required for the latter steps in lipid A synthesis, lpxH, lpxK, WaaA, lpxL and lpxM (Figure 5.1.3).

The coding region surrounding the gene encoding the σ^{54} factor (*rpon*) includes a number of unannotated open reading frames (ORFs), and the *lpxA* gene. This proximity could be coincidental, or it could suggest an additional regulatory connection between σ^{54} expression and lipid A synthesis in *A. aeolicus*.

5.1.3.3e. Aq_087, Aq_1119

It has been previously discussed that NtrC1, hksP1, and the Aq_1119 genes appear to be clustered, similar to the NtrC3, HksP4, dhsU operon (Doucleff, et al., 2005).

Figure 5.1.5. Current model for σ^{54}-dependent regulatory networks in Aquifex aeolicus. Solid black lines indicate published association. Green lines indicate experimentally determined work in this thesis. Dotted lines indicate inferred connections from genomic proximity.



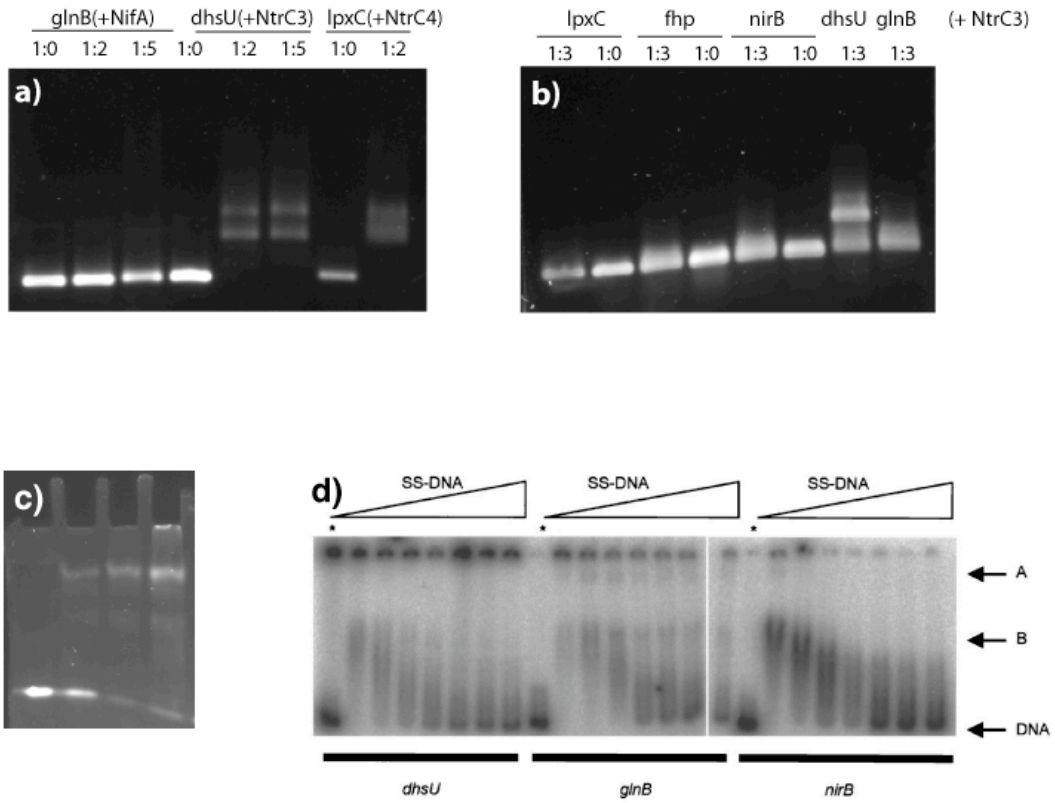
But the genes following and preceding NtrC1/hksP1 in the Aq_1119 region are mostly unannotated. The same is true of the Aq_087 region, which is neighbored by a number of ORFs and two unrelated genes: molybdenum transport protein modC, and an aldolase kdsA. This lack of a functional pattern, interrupted by a series of ORFs, precludes a discussion of these gene functions.

5.1.3.4. Determination of the DNA-binding sites for NtrC3 and NtrC4

Published associations between σ^{54} activators and UAS binding sequences in *A. aeolicus* are limited to NtrC1, which was reported to bind to the *aq_1119* UAS (Doucleff, et al., 2005); and NifA, which was reported to bind to the UAS of *glnB* (Studholme and Wigneshweraraj, et al., 2000). In order to determine other associations between NtrC-like proteins and the putative genes listed above, gel-shift assays were conducted.

Regions spanning 250 base pairs upstream of the *glnB, dhsU, fhp, Aq_087, Aq_1119, nirB*, and *lpxC* genes were amplified by PCR. To determine whether activators bound these UAS regions, gel-shift assays with DNA-binding domains or full-length constructs of NtrC1, NtrC2, NtrC3, and NtrC4 were conducted on each of the 7 potential UAS's. These gels demonstrate associations between NtrC4 and the lpxC-UAS with dual bands in the bound form, suggesting the presence of two binding sites in the region. NtrC4 also bound the 21-base pair NtrC1 high affinity binding site (Kd ~ 0.1 nM) in the *aq_1119* UAS with low affinity (Kd ~ 1 μM) (data not shown). This could be an indicator that NtrC4 and NtrC1 both play a regulatory role on the *aq_1119* gene; however, the high Kd of NtrC4 binding to the *aq_1119* UAS suggests the interaction

Figure 5.1.4. Gel-shift assays with UAS regions and σ^{54} activators. a) glnB, dhsU, and lpxC UAS regions with NifA, NtrC3, and NtrC4, respectively. b) Gel shift indicating that NtrC3 binds the dhsU UAS and no others. DNA concentrations are 1.98, 0.70, 0.88, .36, and 0.66 μM for glnB, dhsU, lpxC, fhp, and nirB, respectively. Gel tracks are labeled with DNA:[monomeric protein] molar ratios. c) 4-20% Tris-Cl gradient native gel showing binding of NtrC3-FL to the dhsU sequence [AGGTGTCatttttGACACCT] at 300 nM concentration. *d) Reprinted from Studholme and Wigneshwereraraj et al., 2000: DNA competition assays demonstrating the binding of NifA to the glnB UAS. Starred tracks contain no NifA. Tracks ramp from 17 to 204 μg/ml salmon sperm nonspecific DNA.*



could be coincidental, and not have physiological relevance. NtrC3 bound to two sites in the *dhsU*-UAS with relatively high affinity (Kd < 300 nM). Aside from these three associations, no other associations were seen (Figure 5.1.4). This includes a lack of association between NifA and *glnB*, as seen by Studholme and Wigneshweraraj, et al., and a lack of any associations for the NtrC2 protein.

Next, I determined the specific 20-base pair NtrC3 and NtrC4 binding sites within the UAS sequences of lpxC and dhsU. To try to identify specific activator binding sites upstream of putative σ^{54}-activated genes, the PALINDROME program (Rice, et al., 2000) was used to search for palindromic sequences in the UAS regions. Although palindromic regions were found in all UASs, only one site with the characteristics of an activator binding site was found ~ 100 bp upstream of the transcription initiation site of the *dhsU* gene. This 21-base pair region [AGGTGTCatttttGACACCT] was purchased

and found to bind to NtrC3-FL in native 20% PAGE gel-shift assays. The second binding site in the *dhsU* region has not been identified.

In the case of NtrC4, binding sites could not be identified based on symmetry and A-tract signatures alone, and the DNA-binding sites were determined by Dnase I footprinting of the UAS of the *lpxC* gene by Dr. Yixin Huo. The Dnase I footprint shows two regions of protection, located –145 and –85 base pairs upstream of the σ^{54} binding site. These binding sites were confirmed by fluorescence anisotropy and gel-shift assays, as discussed in Chapter 2.

5.1.3.5. Implications of NtrC3, NtrC4, NifA regulation

The interaction of NtrC3 with *dhsU* is the first known case of σ^{54} regulation of this gene. A number of examples do exist of σ^{54}-regulation of other sulfur metabolism genes, including the *sfnR* gene in *Pseudomona putida DS1* (Habe, *et al.,* 2007), and *cbl* in *E. coli* (Zimmer, et al., 2000). The interaction of NtrC4 with the *lpxC* gene is also the first time any protein associated with cell wall synthesis has been observed under σ^{54} control. The LpxC enzyme has gained recent interest as a drug target (Raetz and Whitfield, 2002) in Gram-negative bacteria. If this association is also present in pathogenic species, disruption of the σ^{54} pathway would provide an additional pharmaceutical target.

The lack of binding seen between NifA and the *glnB* UAS is unexpected, and contradicts published data. Studholme and Wigneshwereraraj *et al.,* (2000) used mobility shift assays in which 25 nM ^{32}P-labeled glnB-UAS DNA fragments approximately 200 bp in length were bound to 50 nM NifA dimers in the presence of varying quantities of competitor DNA (figure 5.1.4d). The NifA protein was observed to form complexes with *dhsU, glnB* and *nirB* UASs, but only the complex with *glnB* remained upon the addition of competitor DNA. Our assays differed in a number of technical aspects. While the published assays were performed at 25 nM DNA concentration, my assays were performed between 1.9 and 0.36 μM DNA. However, this difference would be expected to enhance both specific and nonspecific binding. Another notable difference is the gel used; Studholme *et al.,* use 4.5% native PAGE gels (the percentage of bis-acrylamide: acrylamide is not specified), where my assays use 1% LMP agarose. My early work with 4% PAGE gels also exhibited a similar smearing affect of the bands as seen in figure 5.1.4d, most likely due to the large size of the DNA (~150,000 kDa) relative to the small size of the protein (100 kDa). However, the difference of gel used does not explain why a faint band is visible in their glnB-NifA gel when we see no interaction. Additional experiments, possibly varying NifA/DNA concentrations and in different buffer conditions, may be necessary to confirm or reject the interaction between NifA and *glnB*. For the purpose of this thesis, the NifA-*glnB* association will be listed in the current model until this inconsistency is explained, or a different association is confirmed.

5.1.3.6. The absence of additional binding associations

It is not clear, from this work, why associations could not be experimentally found between NtrC2 or NifA and any of the σ^{54} activated genes in question. It is also

not known why activators were not found to bind the fhp, Aq_087, and nirB genes. There are a few possible explanations for these observations.

First, activator binding sites may be located outside the 250-bp UAS region. There exist a few unusual examples of systems where the activator binding sites that reside 3' to the σ^{54} promoter, such as flaNQ in *Caulobacter crescentus* (Gober and Shapiro, 1992). It has been shown that activators can activate transcription when transplanted up to 1 kbase upstream of the activation site, or even on a separate, concatenated strand of DNA (Wedel, et al., 1990). Hence, there is no known reason by activators could not be located elsewhere in the genome.

Second, it is possible that more σ^{54}-transcriptional activators exist in *A. aeolicus* than we have identified. Although this would be surprising, given the high level of conservation in σ^{54}-transcriptional activators (Tucker and Sallai, 2007), there is at least one gene in *A. aeolicus* that is a candidate for a "missed" σ^{54}-activator (aq_091m). This gene contains a PAS regulatory domain, an AAA+ ATPase domain and the coding for a DNA-binding domain. However, there is a stop codon in the middle of the AAA+ ATPase domain, which we would expect to render the central domain inactive, as well as prevent the coding of the DNA-binding domain. For this reason, we do not consider it in our analysis. However, a number of AAA+ conserved motifs are intact beyond the stop codon, including the GAFTGA loop and the Walker B motif. Whether this stop codon is a sequencing error, a recent mutation in the genomic DNA used to perform sequencing, or an evolutionary step to disconnect the protein from σ^{54} activation, is not known.

A third possibility, and the most difficult to surmount with current biochemical techniques, is that laboratory conditions are sufficiently different from physiological conditions to preclude binding in our assays. The mechanism by which proteins bind to DNA at high temperatures is not well understood, and it is possible that high temperatures are required for these complexes to form.

5.1.3.7. Current model for σ^{54} gene associations in *A. aeolicus* and future studies

Combining the associations described above, a model of σ^{54}-dependent activation pathways is presented in figure 5.1.5. However, additional data would be valuable. It would be valuable, for example, to try to form open complexes of σ^{54}-RNAP at the σ^{54} sites of putative enhancers to confirm that these regions are indeed σ^{54}-activated. If these sites are confirmed, other techniques could be used to determine whether more binding sites will be determined at higher temperatures. One possibility is repeating the above binding assays at temperatures near the melting point of the DNA in question. To go to higher temperatures, potential short binding sites could be selected from UAS regions and synthesized as DNA-hairpins to reduce the incidence of DNA-melting. If longer regions can be used, GC-rich regions flanking the UAS regions could be inserted to stabilize DNA-annealing during assays.

5.1.3.8. Distances between activator and σ^{54} binding sites

In comparing the activator and σ^{54} binding sites in NifA, NtrC, NtrC1, NtrC3, and NtrC4, it is apparent that different numbers of nucleotides separate the two sites in each system. The distances between the center of the proximal activator-binding site (relative

to the σ^{54} binding sequence) and the –24 base of the σ^{54} consensus sequence are: NifA from *Klebsiella pneumoniae* (118 bp), NtrC from *S. enterica* (80 bp), NtrC1 (49 bp), and NtrC4 from *A. aeolicus* (91 bp). Since only one NtrC3 binding site has been found in *A. aeolicus*, it is not known whether this site is proximal or distal to the σ^{54} binding site. However, like NtrC1's proximal site, its center is located only 49 bases upstream of the –24 σ^{54} element, suggesting that it is the proximal site. This short distance between NtrC1 and NtrC3 activator sites and the σ^{54} binding site is unusual, compared to the observation that activator binding sites are usually 80-120 base pairs upstream of σ^{54} binding sites. The purpose of this closer placement of activator and σ^{54} binding sites in NtrC1 and NtrC3 is not known.

Part 2: Characterization of a histidine kinase: HksP4

5.2.1 Introduction

5.2.1.1. Histidine kinases in TCS pathways: common structural/functional themes

In addition to determining the genes under σ^{54}-control, it is also valuable to determine the environmental signals that activate transcription in σ^{54}-dependent systems. This includes signals that directly activate the GAF regulatory domains of NtrC2 and NifA, as well as the signals that activate the TCS histidine kinases associated with the response regulator domains of NtrC1, NtrC2, and NtrC4. Currently, no activating signals are known for *A. aeolicus* NtrC homologues. The determination of these regulatory signals is difficult, due to the extraordinary range of possible ligands and triggering environmental conditions (Krell, et al., 2010).

As described in the Chapter 1, two-component signal transduction systems (TCSs) consist of a sensor kinase (SK) and a response regulator (RR). The SK is often a multidomain membrane-bound protein with a transmembrane domain, cytosolic autokinase domain and one or more sensor domains. Figure 5.2.1 shows the domain organization of a number of SK's.

PAS domains (for Period, Aryl hydrocarbon receptor nuclear translocator, and Single Minded) are the most common regulatory domains in SKs. The PAS fold is usually ~130 residues long with an α/β fold. However, PAS domains exhibit a great deal of sequence diversity: they contain two conserved regions, averaging only 12% sequence conservation. As a result, PAS domains are sometimes difficult to identify with Markov models. A number of examples exist in the literature where PAS homology was not evident in a protein sequence, but was identified after the three-dimensional structure was solved (Krell, et al., 2010). This low conservation enables the PAS domain to sense a wide array of environmental signals (Gilles-Gonzalez and Gonzalez, 2004).

The prototypical SK autokinase domain can be further divided into two subdomains. The first is the Dimerization/Histidine/phosphorylation domain (DHp). This domain consists of a long α-helix that provides a dimerization determinant for the full-length enzyme, and houses the catalytic histidine that is phosphorylated. The other domain is the catalytic ATP-binding (CA) domain. Both domains are part of the same polypeptide, and interact by forming an interdomain β-sheet.

In the majority of cases (~83%), the SK's are membrane-bound. In cases where extracellular or periplasmic stimuli are identified, such as in PhoQ and LuxQ (Fig 5.2.1), the sensor domain is periplasmic and attached to the cytosolic autokinase domain by a transmembrane region. Only ~ 17% of SK's are fully cytosolic.

5.2.1.2. Heme-containing histidine kinases

Although only a small subset of PAS domains associate with cofactors, these

Figure 5.2.1. Domain organization of a number of SKs. Reproduced from Krell, 2010.

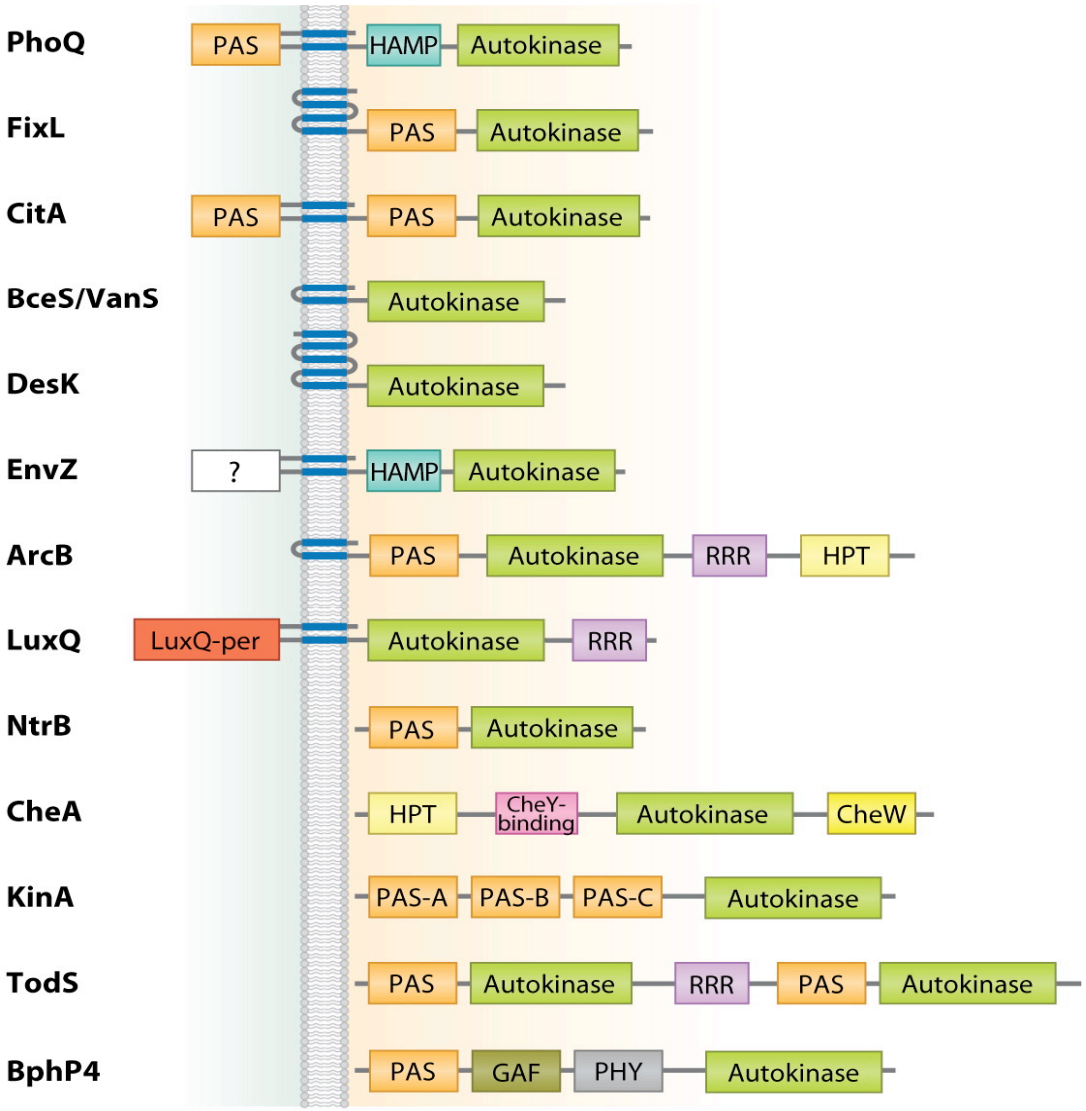
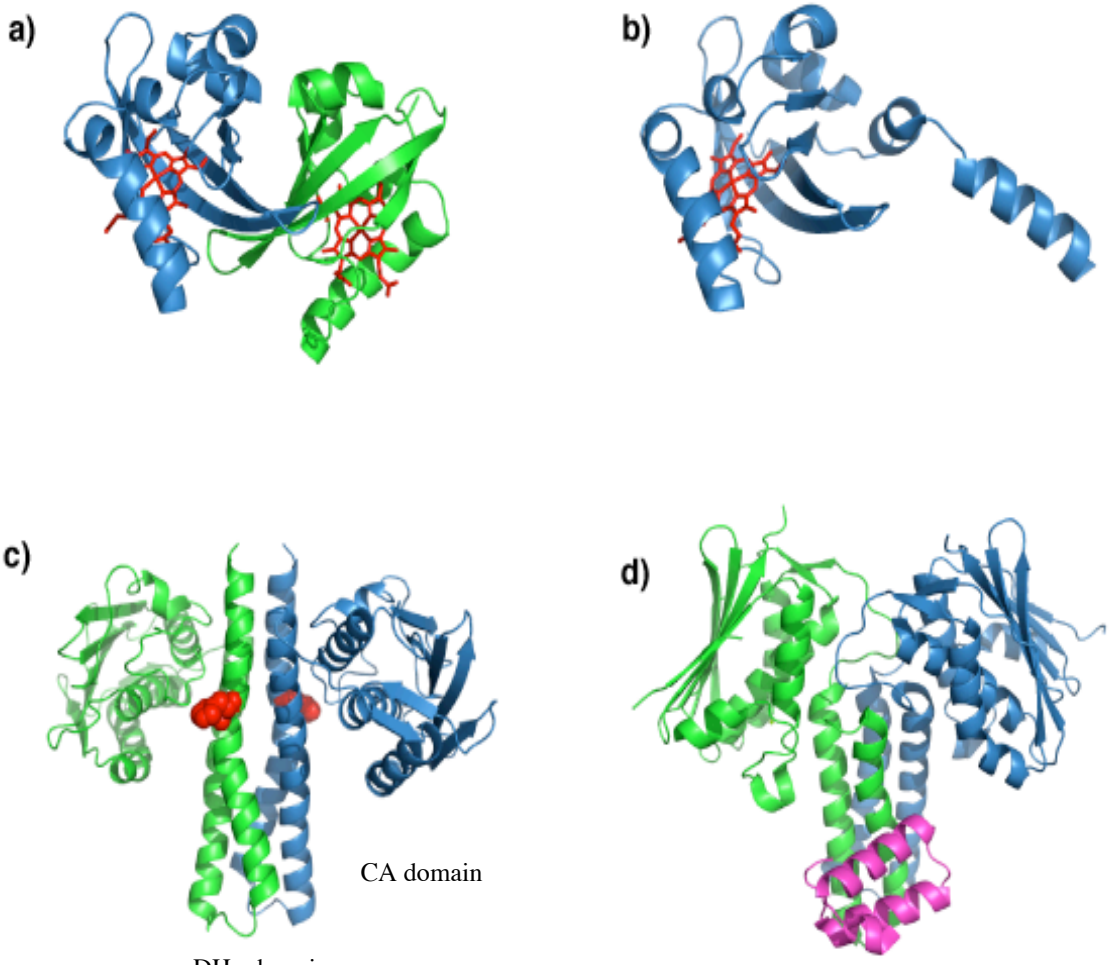


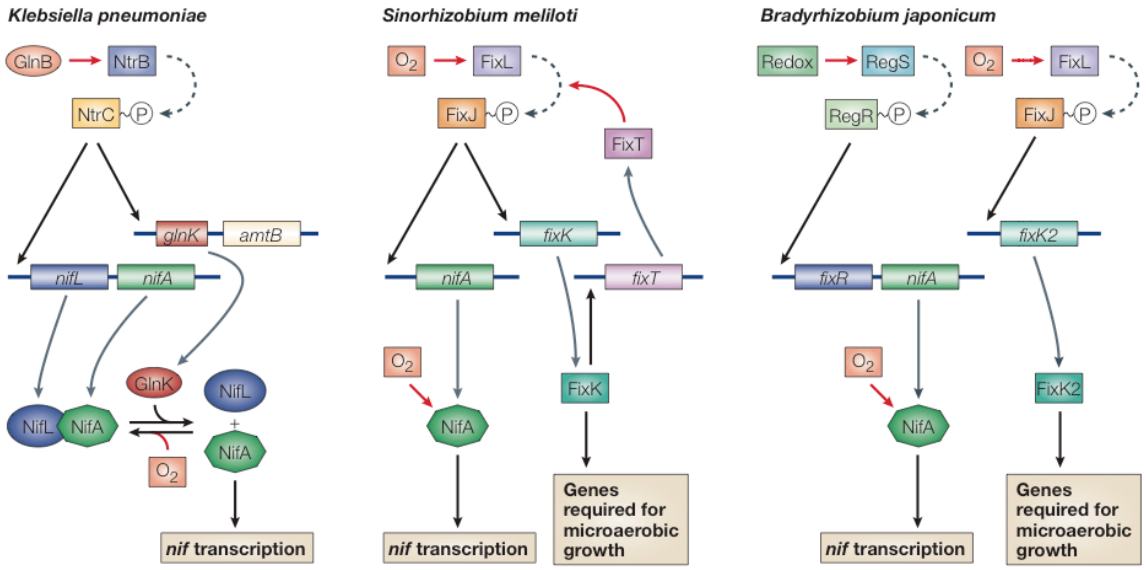
Figure 5.2.2. Structures of histidine kinase PAS and kinase domains. a) bjFixLH, the PAS domain dimer from *Bradyrhizobium japonicum* (Ayers and Moffat, 2008); b) smFixL, the PAS domain from *Sinorhizobium meliloti* (Miyatake, et al., 2000); c) Dimeric DHp and CA domains of HK853 from *Thermatoga Maritima* (Casino et al, 2009). The catalytic histidine (H260) is highlighted in red. d) Dimeric DHp and CA domains of KinB in complex with modulator Sda (Bick, et al., 2009).



proteins constitute some of the best-studied examples in the literature. Sensing of signals such as diatomic gases, light, and redox potential is known to require cofactors such as heme, FMN, or FAD (Gilles-Gonzalez and Gonzalez, 2004). PAS domains containing hemes are commonly sensors for diatomic gases, including O_2, CO, and NO. Although usually associated with prokaryotic organisms, a heme-containing PAS domain was found to sense CO in the mammalian protein NPAS2, implicated in the regulation of circadian rhythms (Dioum, et al., 2002).

The most characterized example of a SK utilizing a PAS domain with a heme cofactor is the FixL protein, an oxygen-sensing histidine kinase that activates its conjugate RR FixJ in *Rhizobia* to activate oxygen-sensitive genes involved in nitrogen fixation. Functionally, the signaling cascades of FixL from rhizobia have some similarities to the signaling cascade of NtrC (Figure 5.2.3). FixL has been shown to bind O_2, CO, and NO with ferrous heme. O_2-binding was found to deactivate the kinase by ~100-fold, whereas O_2 dissociation from the heme restores activity. Binding to CO and NO were also found to deactivate the kinase by 3-fold and 2-fold, respectively. The function of binding these secondary ligands is not known Gilles-Gonzalez and Gonzalez, 2004). However, it may be incidental, since it is uncommon for hemes that bind O_2 to

Figure 5.2.3. Signaling cascades of nitrogen fixation. From Dixon and Kahn, 2004.



fully descriminate against CO and NO (Piantadosi, 2008).

A number of crystal structures of FixL PAS domains have been presented in the literature from *Bradyrhizobium japonicum* (bjFixL) and *Sinorhizobium meliloti* (smFixL) in the unliganded, O_2-bound, CO-bound, and NO-bound states (Ayers and Moffat, 2008; Miyatake, et al., 2000). Structures of bjFixL and smFixL PAS domains are shown in Figure 5.2.2.

5.2.1.3. Introduction to heme biochemistry

To discuss heme sensors more fully, it is useful to briefly outline the basics of heme biochemistry. Heme is a prosthetic group consisting of iron coordinated into a heterocyclic organic ring called a porphyrin. Multiple forms of heme exist, distinguished by different chemical derivatives on the porphyrin ring. The most common form of heme, *heme b* or iron protoporphyrin IX, is shown in Figure 5.2.5. *Heme b* is present in most known metalloproteins, and is often covalently bound to the protein by coordination of the iron atom with the amino group of a histidine side chain (common in globins, cytochromes, nitroporphyrins and peroxidases), the hydroxyl group of a tyrosine side chain (catalases), or the sulfur group of a cysteine (siroheme, nitric oxide synthases, P450 proteins) (Cowan, 1993). Without binding a substrate, these hemes are pentacoordinate; upon binding a diatomic gas, they become hexacoordinate.

The iron atom in hemes can exist in ferric (+3) or ferrous (+2) oxidation states. Ferrous iron is implicated in binding O_2, CO, NO. Ferric heme is implicated in binding OH-, F-, CN-, and H_2S (Kraus, et al., 1990). Heme groups have strong electronic absorption bands that are sensitive to oxidation state, ligation state, and conformation of the chromophores, making UV/Visible spectroscopy a useful tool study them (Nienhaus and Nienhaus, 2005). The strongest feature in the electronic spectrum of heme is the Soret band, which has an extinction coefficient ~ 100,000 $M^{-1}cm^{-1}$ and is centered around ~420-440 nm. There are also a few minor bands, denoted Q bands (alpha and beta) in the 500-600 nm region. An example of the Soret band shifts observed in the FixL-PAS domain binding to O_2, CO, and NO is shown in Figure 5.2.7.

5.2.1.4. Characterization of a histidine kinase in *A. aeolicus*: HksP4

HksP4 is one of four histidine kinases in *A. aeolicus* with unknown functions (other histidine kinases include *phoB, gyrB, mutL*). As described previously, NtrC1-HksP1 and NtrC3-Hksp4 appear to be associated TCS pairs (Doucleff, 2005). The domain architectures of *A. aeolicus* histidine kinases determined by BLASTp is shown in Figure 5.2.6. This search finds that HksP4 is a histidine kinase with a PAS domain with a heme-binding motif, and a DHp-CA kinase domain with an ATP and Mg^{2+} binding site. It is predicted to have no transmembrane regions by the TMHMM server (Tusnady and Simon, 1998). The fact that HksP4 is predicted to bind heme suggests that the determination of the kinase ligand may be restricted to diatomic gases, as seen in the FixL system. The fact that HksP4 is predicted to be a fully soluble protein suggests that its expression and purification may be straightforward. Therefore, we decided that HksP4 was a suitable system to characterize further, with the goal of elucidating a full TCS regulatory pathway in *A. aeolicus*, from the activating signal to the gene activated.

Unfortunately, although HksP4 can be considered a FixL-like heme-binding PAS-DHp-CA SK, the PAS domain of HksP4 and FixL share no detectable sequence identity. As a result, a structural understanding of the HksP4-PAS domain cannot be derived from studies on FixL. However, a number of very recent structures have emerged of histidine kinases with DHp and CA domains homologous to HksP4, many of them in complex

Figure 5.2.4. UV spectra showing Soret band shifts for the O_2-bound and deoxygenated states of the B. japonicum FixL PAS domain (bjFixLH). Adapted from Liebl, et al., 2002.

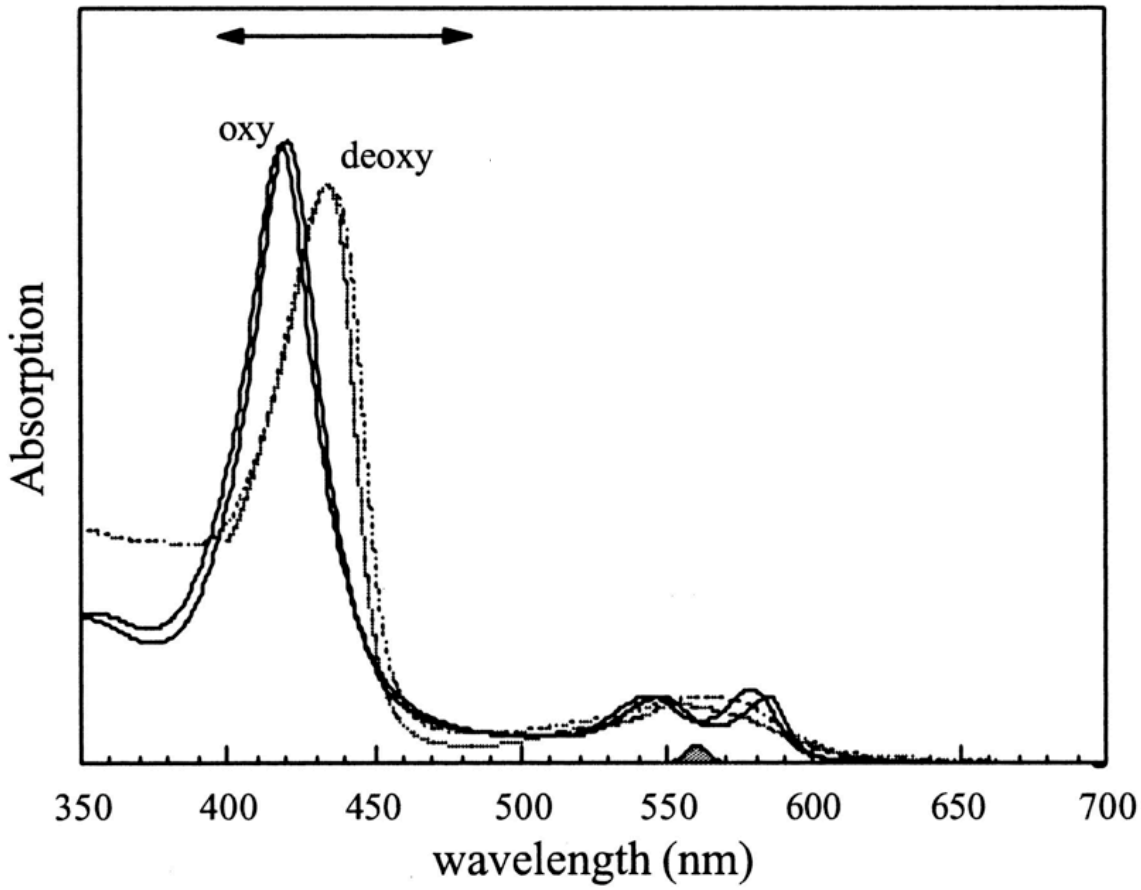
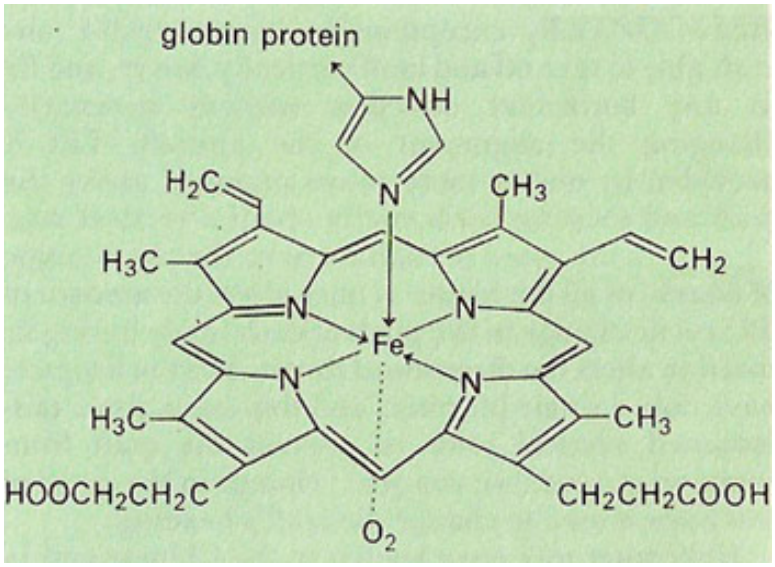


Figure 5.2.5. The structure heme b, also known as Iron protoporphyrin IX. From http://www.daviddarling.info/encyclopedia/H/heme.html.



with their response regulators. This includes the structures of HK835 from *Thermatoga maritima* (Casino, *et al*., 2009) and KinB from *Geobacillus stearothermophilus* (Bick, *et al*., 2009). The histidine kinase ThkA, which contains a coiled coil, GAF, PAS, and CA domain, was truncated to the PAS-ATPase core and its struture was solved in complex with its response regulator (Yamamoto, *et al*., 2009). However, due to the inherently low sequence similarity of PAS domains (Pandini, 2005), the ThkA PAS domain, like the FixL PAS domain, shares no detectable homology with the PAS domain of HksP4. at the time of writing this thesis, no HksP4-like PAS domain structure was deposited in the PDB database. As such, no current structural information is available for a full-length homologue of HksP4.

5.2.2. Materials and Methods

5.2.2.1. Cloning

The full-length HksP4 construct was amplified by PCR from *A. aeolicus* genomic DNA with NdeI and BamHI cut sites. The PCR product was cut with NdeI and BamHI enzymes, and ligated into a pskB2 plasmid with an N-terminal 6xHis tag with T4 DNA ligase (New England Biolabs). Plasmids were transformed into Top10 highly competent cells (Invitrogen), and plasmid was isolated and its sequence confirmed by DNA-sequencing at the UC Berkeley Sequencing Facility. The HksP4 PAS domain construct (1-116) was produced by introducing a stop codon at residue 116 using Quikchange (Stratagene).

5.2.2.2. Protein expression of HksP4-FL and HksP4-PAS

Protein was resuspended in buffer A: 50 mM Tris-HCl, 200 mM NaCl, pH 8.0, 1 mM TCEP. Resuspended cells were supplemented with protease cocktail inhibitor (Roche) and sonicated for 6x30s intervals at high power and 4ºC. Lysate was ultracentrifuged at 30,000 rpm for 30 minutes, and supernatant was purified by Ni-NTA resin (Qiagen) as described in chapter 4. Samples were eluted into 50 mM Tris-HCl buffer, 300 mM NaCl, and 400 mM imidazole, pH 8.0, and dialyzed into Buffer A overnight (with 1 mM DTT in place of 1 mM TCEP). Samples were concentrated in Amicon centrifuge concentrators.

5.2.2.3. Activity assays on HksP4

In collaboration with Dr. Brian Smith of the Marletta lab, two forms of ATP-activity assays were performed on HksP4. These assays both detect the ability of the kinase to autophosphorylate, and its ability to transfer the phosphate group to the response regulator. Both assays are described in Carlson et al., 2000: ^{32}P ATPase assays, and ATP-γS assays (Carlson, 2010).

Preparation of HksP4 liganded and unliganded states:

Anaerobic: Protein was concentrated to 100 μM, and deoxygenated by incubation with 10 mM sodium dithionite for 40 minutes in a glove box with [O_2] at 10-13 ppm. To

remove the dithionite, the sample was buffer exchanged into buffer A with a P-10 buffer exchange column with G-25 resin (Amersham) that had been degassed by equilibration with buffer containing 50 mM sodium dithionite and washed with 50 mL degassed buffer A.

O_2-bound: anaerobic sample in buffer A was exposed to air for 5 minutes.

CO-bound/H_2S-bound: anaerobic sample in buffer A was placed in a septum-capped vessel under CO or H_2S gas for 10 minutes or 1 hour, respectively.

NO-bound: anaerobic sample was mixed with a small amount (~ 1 μM) of diethylNONOate, a compound that spontaneously forms NO in solution. Sample was incubated for 5 minutes before analysis.

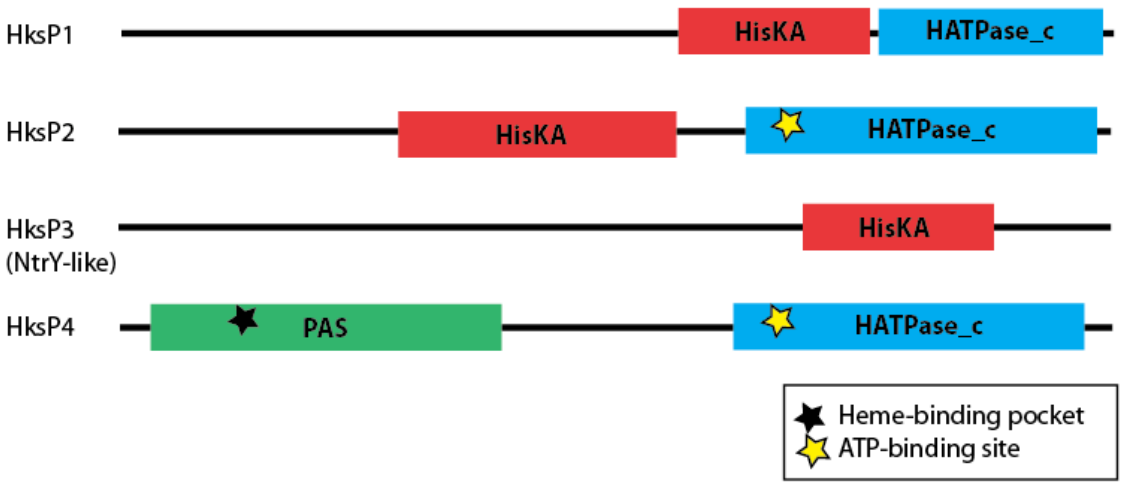
Ferric (oxidized): HksP4 in buffer A was incubated with 1 mM KFeCN for 20 minutes, then buffer-exchanged back into buffer A with PD-10 column to remove the KFeCN.

UV-vis spectra from 200 to 800 nm were taken on each sample to assess binding in a septum-sealed anaerobic cuvette at 25ºC on a Cary UV spectophotometer.

ATP-γS assays: Assays were performed as described in Carlson, et al., 2010. 20 μL aliquots were prepared that contained 5 μM HksP4-FL in the anaerobic, O_2-bound, NO-bound and CO-bound states, 500 μM ATP-γS, and 10 mM $MgCl_2$ in buffer A. To search for phosphotransferase activity, some samples also contained 5 μM NtrC3-FL. Aliquots were quenched by addition of 5 μL of 0.5 M EDTA. Paranitrobenzylmethlate (PNBM), a compound that alkylates thiophosphate groups, was added to a final concentration of 1 mM and alkylation reaction was allowed to proceed for 1.5 hours at room temperature. 8 μL of 4x SDS-PAGE protein loading buffer was added, and 10 μL of each aliquot was run on a 20% SDS-PAGE glycine gel for 1 hour at 180 V. Proteins were transferred from gels to nitrocellulose membranes (Whatman) and blocked for 1 hour with 5% nonfat dry milk (Carnation) in phosphate-buffered saline (PBST), pH 8.0, containing 0.5% Tween20. Primary antibody specific for the alkylated thiophosphate (Epitomics antibody 51-8) was added at 1:5000 in 5% nonfat dry milk in PBST and incubated with the blot overnight at 4ºC. The blot was washed three times in10 minute aliquots with PBST. Secondary antibody (HRP, Pierce) was added at 1:1000 in 5% nonfat dry milk in PBST for 1 hour. The blot was then washed 3x10 minutes with PBST and developed by the addition of SuperSignal West Femto Maximum Sensitivity Substrated (Pierce), and imaged with a BioRad Chemidox XRS chemiluminescence imager.

^{32}P-ATPase assays: 20 μL aliquots were prepared that contained 5 μM HksP4-FL, 500 μM ATP supplemented with 10 μCi ^{32}P-ATP per aliquot, 10 mM $MgCl_2$, and in samples that contained activator, 5 μM NtrC3-FL. Samples were incubated at room temperature in microcentrifuge tubes sealed in septum-capped vials to maintain anaerobicity. Samples were quenched with the addition of 5 μL 4X-concentrated SDS-protein loading buffer, and run on a 10-20% Tris-Cl gradient PAGE gel for 2.5 hours at 150 V. Gels were dried overnight and imaged on a phosphorimager plate with a Typhoon (Molecular Dynamics).

Figure 5.2.6. BLAST domain prediction in the four A. aeolicus histidine kinases.



5.2.3. Results and Discussion

5.2.3.1. HksP4 contains heme, and binds O_2, NO, and CO, not H_2S

Two constructs of HksP4 were cloned: full-length (HksP4-FL; 1-339) and the PAS domain (HksP4-PAS: 1 to 116). Both proteins express with high yield, averaging 30 mg/L TB media. Both proteins are also very soluble: HksP4-PAS could be concentrated to 30 mg/mL, and HksP4-FL could be concentrated to 15 mg/mL, although both constructs precipitate after a 1-2 days at these concentrations. HksP4-PAS and HksP4-FL were also both found to bind heme. Both constructs have a distinctly brown color upon purification, and the UV spectrum of the protein from 200 to 800 nm include a signature Soret band. However, over the course of protein purification HksP4 loses its heme, possibly due to interaction with imidazole in the Nickel-column elution buffer. Heme-binding could be reconstituted by titrating hemin directly into the purified HksP4 solution in a 1:1 hemin:protein molar ratio, and removing the excess hemin with a PD-10 buffer exchange column. (Figure 5.2.7).

Although the three diatomic gases best known to bind to hemes are O_2, CO, and NO, H_2S is also a gaseous ligand of ferric heme. Although no PAS-associated heme has been studied binding H_2S, it has long been known that H_2S binds to heme with toxic effects in humans (Keilin, 1933). Recent studies have generated new interest in H_2S-heme interactions, since new evidence suggests that H_2S serves a beneficial role as a neurotransmitter in mammals, regulating vascular tone, neurotransmission, and insulin secretion in humans (Lowicka and Beltowski, 2007). Additionally, since the *dhsU* gene is responsible for sulfur metabolism, possibly using sulfide as a substrate, one possible hypothesis is that H_2S activates HksP4, and subsequently the *dhsU* gene. For these reasons, H_2S was also included in our analysis.

UV spectra of HksP4-PAS in the anaerobic state and in the presence of O_2, CO, NO, and are shown in Figure 5.2.7. It is evident from the shifts of the Soret band that

Figure 5.2.7. UV-vis spectra of HksP4-FL binding to O_2, CO, and NO. Curves are normalized to the same concentration. Peak centers, in nm, are indicated above each curve.

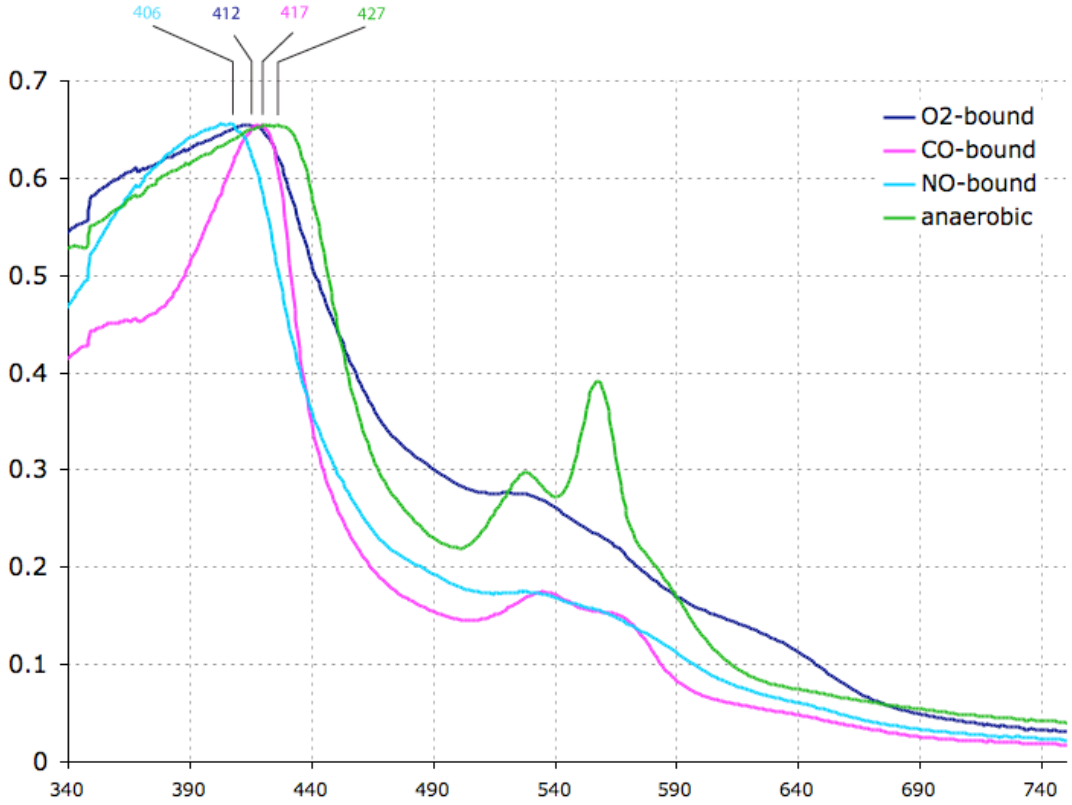
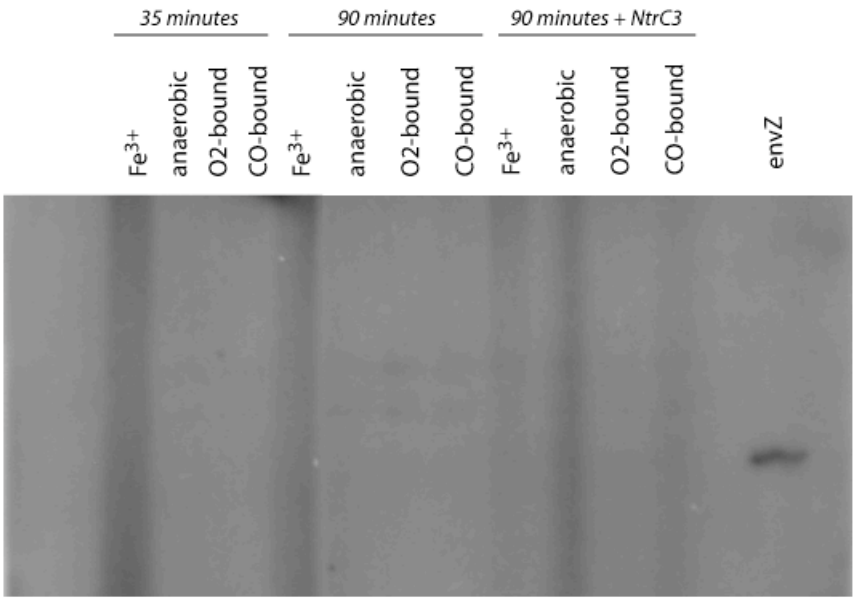


Figure 5.2.8. ATPγS assay of HksP4-FL while anaerobic, and binding to O_2, CO, NO, and H_2S. The EnvZ kinase was included as a positive control.



HksP4 binds O_2, CO, and NO. No shift was observed in binding to H_2S (data not shown).

5.2.3.2. Activity assays: anaerobic, O_2-, and CO-bound HksP4 do not exhibit ATPase activity.

Activity assays, including both ^{32}P-ATPase assays and ATP-γS assays, show essentially no ATPase activity in HksP4 in the anaerobic ferrous, unliganded ferric, and ferrous-O_2-bound, CO-bound, and NO-bound states. EnvZ, a histidine kinase from *E. coli*, was included as a positive control (Figure 5.2.8).

There are a number of possible causes that could explain the lack of activity. First, as is unavoidable in themophilic proteins, it is possible that the experimental conditions of the assay, including 200 mM NaCl and room-temperature incubation, are sufficiently different from physiological conditions that they preclude activity. It is also possible that the heme-protein interaction is not native. Since HksP4 was expressed in *E. coli* cells with supplemented amino-levulinic acid, a heme b precursor, less common forms of heme that may be present in *A. aeolicus* might not be expressed in sufficient quantities to form the native protein-heme complex. The observation that HksP4 constructs lose the heme over the course of purification would be consistent with this hypothesis.

Second, it is possible that autophosphorylation by the kinase requires an additional protein or chemical modulator. This has been observed in the case of NtrB, the soluble histidine kinase with a PAS domain that activates NtrC in *E. coli*. Unlike most SKs, NtrB does not sense a stimulus directly, but is regulated by an additional sensor protein PII, encoded by the *glnB* gene, in the pathways described in Figure 5.2.3. Although the heme-binding PAS domain in HksP4 suggests that it does have sensor activity, it may require an additional modulator to enable autophosphorylation.

5.2.4. Concluding remarks

A main adaptation to autotrophy is the elevated use of metalloenzymes to perform processes ranging from nitrogen and carbon fixation to respiration (Deckert, et al., 1998). Extremophiles have a particularly challenging task of using metal chemistry efficiently in extreme environmental conditions without poisoning themselves.

A few possible speculations about σ^{54} regulation in *A. aeolicus* would be consistent with this difficulty. The placement of *nirB, fhp,* and *lpxC* genes under σ^{54} control suggests sensitive regulation of nitrogen assimilation and lipoprotein synthesis, depending on environmental conditions. The regulation of *dhsU* by a histidine kinase that binds oxgen suggests that *A. aeolicus* may share similarities with other vent organisms that use hydrogen sulfide as a reductant to oxidize oxygen to fuel respiration (Stetter, 2006). It is also possible that, like the FixL protein, HksP4 binds to O_2 as an O_2-sensor to prevent the expression of O_2-sensitive metalloenzymes in sulfide oxidation and other downstream pathways.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66(Pt 2), 213-21.

Ababou, A., Rostkova, E., Mistry, S., Le Masurier, C., Gautel, M., Pfuhl, M. (2008) Myosin Binding Protein C Positioned to Play a Key Role in Regulation of Muscle Contraction: Structure and Interactions of Domain C1. J. Mol. Biol. 384, 615-630.

Amouyal, M. (2005) Gene regulation at-a-distance in E. coli: new insights. C R Biol. 328, 1-9.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Lakshminarayan, M.I. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. FEMS Microbiol Rev. 29, 231-262.

Arieli, B., Y. Shahak, D. Taglicht, G. Hauska, and E. Padan. 1994. Purification and characterization of sulfide-quinone reductase (SQR), a novel enzyme driving anoxygenic photosynthesis in Oscillatoria limnetica. J. Biol. Chem.

Arrowsmith, C.H., Pachter, R., Altman, R.B., Iyer, S.B., Jardetzky, O. (1990) Sequence-specific 1H NMR assignments and secondary structure in solution of Escherichia coli trp repressor Biochemistry, 29, 6332-6341.

Atkinson, M.R., Narinporn, P., Ninfa, A.J. (2002) Governor of the glnAp2 promoter of Escherichia coli. Mol. Microbiol. 46,1247-1257.

Ayala, I., Sounier, R., Usé, N., Gans, P. Boisbouvier, J. (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. K Biomol. NMR. 43, 111-119.

Ayers, R. A., Moffat, K. (2008) Changes in quaternary structure in the signaling mechanisms of PAS domains. Biochemistry 47, 12078-12086.

Barb, A.W., Zhou, P. (2008) Mechanism and inhibition of LpxC: an essential zinc-dependent deacetylase of bacterial lipid A synthesis. Curr Pharm Biotechnol. 9, 9-15.

Barne, K.A., Bown, J.A., Busby, S.J., Minchin, S.D. (1997) Region 2.5 of the Escherichia coli RNA polymerase s70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. EMBO J. 16, 4034-4040.

Barrios, H., Valderrama, B., Morett, E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. Nuc. Ac. res. 27, 4305-13.

Batchelor, J.D., Doucleff, M., Lee, C.J., Matsubara, K., De Carlo, S., Heideker, J., Lamers, M.H., Pelton, J.G., Wemmer, D.E. (2008) Structure and regulatory mechanism of *Aquifex aeolicus* NtrC4: variability and evolution in bacterial transcriptional regulation. J Mol Biol. 384, 1058-75.

Beck, J.F., Bradbury, C.M., Connors, A.J., Donini, J.C. (1981) Nitrite as antidote for acute hydrogen sulfide intoxication? Am Ind Hyg Assoc J., 42, 805-9.

Beier, D., Gross, R. (2008) Regulation of bacterial virulence by two-component systems. Curr. Op. Microbiol. 9, 143-152.

Betermier, M., Galas, D.J., Chandler, M. (1994) Interaction of Fis protein with DNA: Bending and specificity of binding. Biochimie. 76, 958-967.
Beveridge, D.L., Dixit, S.B., Barreiro G, Thayer KM. (2004) Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. Biopolymers. 73, 380-403.
Bick, M.J., Lamour, V., Rajashankar, K.R., Gordiyenko, Y., Robinson, C.V., Darst, S.A. (2009) How to Switch Off a Histidine Kinase: Crystal Structure of Geobacillus stearothermophilus KinB with the inhibitor Sda. JMB 386, 163-177.
Bokal, A., Ross, W., Gaal T., Johnson R., Gourse, R. (1997) Molecular anatomy of a transcription activation patch: FIS-RNA polymerase interactions at the Escherichia coli rrnB P1 promoter. EMBO J. 16, 154-162.
Bose, D., Pape, T., Burrows, P.C., Rappas, M., Wigneshweraraj, S.R., Buck, M., Zhang, X. (2008) Organization of an activator-bound RNA polymerase holoenzyme. Mol. Cell 32, 337-46.
Brahmachary, P., Dashti, M.G., Olso, J.W., Hoover, T.R. (2004) Helicobacter pylori FlgR is an enhancer-independent activator of sigma54-RNA polymerase holoenzyme. J Bacteriol. 186, 4535-42.
Buck, M. Gallegos, M.-T., Studholme, D.J., Guo, Y., Gralla, J.D. (2000) The Bacterial Enhancer-Dependent sigma 54 (sigma N) Transcription Factor. J Bacteriology. 182, 4129-4136.
Burgess, R.R. (1969) Separation and characterization of the subunits of ribonucleic acid polymerase, J. Biol. Chem. 244, pp. 6168–6176.
Burke, J.R., Deshong, A.J., Pelton, J.G., Rubin S.M. (2010) Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. J. Biol. Chem. 285, 16286-93.
Cai, M., Huang, Y., Kazuyasu, S., Clore, M.G., Gronenborn, A.M., Craigie, R. (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. J Biomol NMR. 11, 97-102.
Campbell, A.E., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., Darst, S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell. 9, 527-539.
Campbell, A.E., Westblade, L.F., Darst, S.A. (2008) Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. Curr. Op. Microbiol. 11, 121-127.
Carlson, H.K., Plate, L., Price, M., Allen, J.J., Shokat, K.M., Marletta, M.A. (2010) Use of a semisynthetic epitope to probe histidine kinase activity and regulation. Anal. Biochem. 397, 139-143.
Casino, P., Rubio, V., Marina, A. (2009) Structural Insight into Partner specificity and phosphoryl Transfer in Two-Component Signal Transduction. Cell 139, 325-336.
Chen, P., Reitzer, L. (1995) Active Contribution of Two Domains to Cooperative DNA Binding of the Enhancer-Binding Protein Nitrogen Regulator I (NtrC) of Escherichia coli: Stimulation by Pohsphorylation and the Binding of ATP. J. Bac. 177, 2490-2496.

Cheng, Y., Wie-Zen Y., Johnson R., Yuan H. (2000) Structural Analysis of the Transcriptional Activation Region on Fis: Crystal Structures of Six Fis Mutants with Different Activation Properties. J. Mol. Biol. 302, 1139-1151.

Cheng, Y.S., Yang, W.Z., Johnson, R.C., Yuan, H.S. (2000) Structural Analysis of the Transcriptional Activation Region on Fis: Crystal Structures of Six Fis Mutants with Different Activation Properties. J. Mol. Biol. 302, 1139-1151.

Cheung, J., Hendrickson, W.A. (2010) Sensor domains of two-component regulatory systems. Curr. Op. Microbiol. 13, 116-123.

Cho, B.K., Knight, E.M., Barrett, C.L., Palsson, B.O. (2008) Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. Genome Res. 18, 900-910.

Clore, G.M. (2008) Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. Mol biosyst. 5, 1058-69.

Cock, P.J., Whitworth, D.E. (2007) Evolution of prokaryotic two-component system signaling pathways: gene fusions and fissions. Mol. Biol. Evol. 24, 23555-57.

Cole C, Barber JD & Barton GJ. Nucleic Acids Res. 2008.

Cowan, J.A. (1993) Inorganic Biochemistry: An Introduction. VCH Publishers, New York.

Craven, C.J., Al-Owais, M., Parker, M.J. (2007) A systematic analysis of backbone amide assignments achieved via combinatorial selective labelling of amino acids. J Biomol NMR. 38, 151-159.

Da Silva Neto JF, Koide, T., Gomes, S.L., Marques, M.V. (2010) Global gene expression under nitrogen starvation in Xylella fastidiosa: contribution of the sigma54 regulon. BMC Microbiol. 10, 231.

De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E., Nixon, B.T. (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. Genes Dev. 20, 1485-95.

De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E., Nixon, B.T. (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. Genes Dev. 20: 1485-1495.

de Vries, S.J., van Dijk, A.D.J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, A., Bonvin, A.M.J.J. (2007) HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins: Struc. Funct. & Bioinformatics. 69, 726-733.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J. and Swanson, R.V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392, 353-358.

Diekmann, S. (1987) Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. Nuc. Ac. Res. 15, 247-265.

Dioum, E., L., Rutter, J., Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.-A., Mcknight, S.L. (2002) NPAS2: A gas-responsive transcription factor. Science 298, 2385-2387.

Dittmer, J., Bodenhausen, G. (2004) Evidence for slow motion in proteins by multiple refocusing of heteronuclear nitrogen/proton multiple quantum coherences in NMR. J Am. Chem. Soc. 125, 1314-5.
Dixon, R., Kahn, D. (2004) Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol. 2, 621-631.
Dombrecht, B., Marchal, K., Vanderleyde, J., Michiels, J. (2002) Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales. Genome Biology 3, 0076.1-0076.11.
Dominguez, C., Boelens, R., Bonvin, A.J. (2003). HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. J. Am. Chem. Soc. 125, 1731-1737.
Dorsey, N.E. "Properties of Ordinary Water-Substance." New York (1940), p. 184.
Doucleff, M.C, Malak, L.T., Pelton, J.G., Wemmer, D.E. (2005) The C-terminal RpoN domain of sigma54 forms an unpredicted helix-turn-helix motif similar to domains of sigma70. J Biol Chem. 280, 41530-6.
Doucleff, M.C., Chen, B., Maris, A.E., Wemmer, D.E., Kondrashkina, E., Nixon, B.T. (2005) Negative regulation of AAA + ATPase assembly by two component receiver domains: a transcription activation mechanism that is conserved in mesophilic and extremely hyperthermophilic bacteria. J Mol Biol. 353, 242-55.
Doucleff, M.C., Pelton, J.G., Lee, P.S., Nixon, B.T., Wemmer, D.E. (2007) Structural basis of DNA recognition by the alternative sigma-factor, sigma54. J Mol Biol. 369, 1070-8.
Emsley, P. & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.
Feldman-Cohen, L.S., Shao, Y., Meinhold, D., Miller, C., Colón, W., Osuna, R. (2006) Common and variable contributions of Fis residues to high-affinity binding at different DNA sequences. J Bacteriol. 188, 2081-95.
Feng, L., Lee, H-S., Prestegard, J.H. (2007) NMR resonance assignments for sparsely 15N labeled proteins. J. Biomol. NMR. 38, 213-219.
Fiaux, J., Bertelsen, E.N., Horwich, A.L., Wuthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex. Nature. 418, 207-11.
Frohlich, K.S., Vogel, J. (2009) Activation of gene expression by small RNA. Curr Opin Microbiol. 12, 674-82.
Gallegos, M.T. & Buck, M. (1999). Sequences in sigmaN determining holoenzyme formation and properties. J. Mol.Biol, 288, 539-553.
Gilles-Gonzalez, M., Gonzalez, G. (2004) Signal transduction by heme-containing PAS-domain proteins. J Appl Physiol 96, 774-783.
Gober, J.W., Shapiro, L. (1992) A developmentally regulated Caulobacter flagellar promoter is activated by 3' enhancer and IHF binding elements. Mol. Biol. Cell. 8, 913-26.
Gong, W., Hao, B., Mansy, S., Gonzalez, G., Gilles-Gonzalez, M.A., Chan, M.K. (1998) Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. Proc Natl Acad Sci. 95, 15177-15182.
Goto, N.K., Kay, L.E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr Op. Struct. Biol. 10, 585-592.

Gotoh, Y., Eguchi, Y., Watanabe, T., Okamoto, S., Doi, A., Utsumi, R. (2010) Two-component signal transduction as potential drug targets in pathogenic bacteria. Curr. Op. Microbiol. 13, 232-239. O_2

Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J., Busby, S.J. (2005) Studies of the distribution of Escherichia coli camp-receptor protein and RNA polymerase along the E. coli chromosome. Proc. Natl. Acad. Sci. 102, 17693-17698.

Grzesiek, S., Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. J. Magn. Reson. 96, 432-440.

Guagliardi, A., Napoli, A., Rossi, M., Ciaramella, M. (1997) Annealing of complementary DNA strands above the melting point of the duplex promoted by an archaeal protein. J. Mol. Biol. 267, 841-848.

Guiral, M., Prunetti, L., Lignon, S. Lebrun, R., Moinier, D., Giudici-Orticoni, M.-T. (2009) New insights into the Respiratory Chains of the Chemolithoautotrophic and Hyperthermophilic Bacterium *Aquifex aeolicus*. J. Proteome Res. 8, 1717-1730.

Habe, H., Kouzuma, A., Endoh, T., Omori, T., Yamane, H., Nojiri, H. (2007) Transcriptional regulation of the sulfate-starvation-induced gene sfnA by a {sigma}54-dependent activator of Pseudomonas putida. Microbiol. 153, 3091-3098.

Haran, T.E., Mohanty, U. (2009) The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys. 42, 41-81.

Harvey, S.C., Dlakic, M., Griffith, J., Harrington, R., Park, K., Sprous, D., Zacharias, W. (1995) What is the basis of sequence-directed curvature in DNAs containing A tracts? J. Biomol. Struct. Dyn. 13, 301-7.

Hastings, C.A., Lee, S.Y., Cho, H.S., Yan, D., Kustu, S., Wemmer, D.E. (2003) High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. Biochemistry. 42, 9081-90.

Hengen, P.N., Bartram, S.L., Stewart, L.E., Schneider, T.D. (1997) Information analysis of Fis binding sites. Nucleic Acids Res. 25, 4994-5002.

Hertz, G.Z., Stormo, G.D. (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics, 15, 563-577

Hirschman, J., P. K. Wong, K. Sei, J. Keener, and S. Kustu. 1985. Products of nitrogen regulatory genes ntrA and ntrC of enteric bacteria activate glnA transcription in vitro: evidence that the ntrA product is a sigma factor. Proc. Natl. Acad. Sci. USA 82:7525-7529

Hong, E., Doucleff, M., & Wemmer, D.E. (2009) Structure of the RNA Polymerase Core-Binding Domain of sigma54 Reveals a Likely Conformational Fracture Point. J. Mol. Biol (2009) 390, 70-82.

Hsu, J.L., Martinis, S.A. (2007) A Flexible peptide tether controls accessibility of a unique C-terminal RNA-binding domain in leucyl-tRNA synthetases. J Mol Biol. 376, 482-91.

Hunt, T. P., and B. Magasanik. 1985. Transcription of glnA by purified Escherichia coli components: core RNA polymerase and the products of glnF, glnG, and glnL. Proc. Natl. Acad. Sci. USA 82:8453-8457

Isono K, Shimizu M, Yoshimoto K, Niwa Y, Satoh K, Yokota A, Kobayashi H Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of □√70 factors of bacterial RNA polymerases in Arabidopsis thaliana. Proc Natl Acad Sci USA 1997 , 94:14948-14953

Jackson, E.N., Yanofsky, C. (1974) Localization of two functions of the phosphoribosyl anthranilate transferase of Escherichia coli to distinct regions of the polypeptide chain. J Bacteriol. 117, 502-8.

Jiang, P., Ninfa, A.J. (2009) Alpha-ketoglutarate controls the ability of the Escherichia coli PII signal transduction protein to regulate the activities of NRII (NrB but does not control the binding of PII to NRII. Biochemistry. 458, 11514-21.

Kalodimos, C.G., Folkers, H.E., Boelens, R., Kaptein, R. (2001) Proc. Natl. Acad. Sci. 22, 6039-44

Kay, L.E. (2004) NMR studies of protein structure and dynamics. J. Mag. Res. 173, 193-207.

Kay, L.E., Xu, G.Y., Singer, A.U., Mugandiram, D.R., Forman-Kay, J.D. (1993) A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H_2O Samples of Proteins. J. Magn. Reson. B 101, 333-337.

Kazmierczak, M.K., Wiedmann, M., Boor, K.J. (2005) Alternative Sigma Factors and Their Roles in Bacterial Virulence. Microbiol. Mol. Biol. Rev. 69, 527-543.

Keilin, D. (1933) On the combination of methaemoglobin with H~S. D. Proc. R. Soc. London, Ser. B: Biol. Sci., 133, 393.

Kelly, M.J.S., Krieger, C., Ball, L.J., Yu, Y., Richter, G., Schmieder, P., Bacher, A., Oschkinat, H. (1999) Application of amino acid type-specific 1H- and 14N-labeling in a 2H-15N-labeled background to a 47 kDa homodimer: Potential for NMR structure determination of large proteins. J. Biomol. NMR. 14, 79-83.

Kercher, M.A., Lu, P., Lewis, M. (1997) Lac repressor-operator complex. Curr. Op. Struc. Biol. 7, 76-85.

Kern, D., Volkman, B.F., Luginbuhl, P., Nohaile, M.J., Kustu, S., Wemmer, D.E. (1997) Structure of a transiently phosphorylated switch in bacterial signal transduction. Nature. 402, 894-898.

Klinge, S., Núñez-Ramírez, R., Llorca, O., Pellegrini, L. (2009) 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. EMBO J. 28, 1978-1987.

Klose, K.E., North, A.K., Stedman, K.M., Kustu, S. (1994) The major dimerization determinants of the nitrogen regulatory protein NTRC from enteric bacteria lie in its carboxy-terminal domain. J Mol Biol. 241, 233-45.

Knauss, R., Fleischer, G., Gruender, W., Kaerger, J., Werner, A. (1996) Pulsed field gradient NMR and nuclear magnetic relaxation studies of water mobility in hydrated collagen II. Magn Reson Med. 36, 241-8.

Kostanjevecki V, Brigé A, Meyer T, Cusanovich M, Guisez Y, Beeumen J. J bacteriol. (2000) A Membrane-Bound Flavocytochrome c-Sulfide Dehydrogenase from the Purple Phototrophic Sulfur Bacterium Ectothiorhodospira vacuolata. J. Bacteriol. 182, 3097-3103

Kostrewa, D., Granzin, J., Stock, D., Choe, H.W., Labahn, J., Saenger, W. (1992) Crystal structure of the factor for inversion stimulation FIS at 2.0 A resolution. J. Mol. Biol. 226, 209-226

Kraus, D. W., Wittenberg, J. B., Jing-Fen, L. and Peisach, (1990) Hemoglobins of the Lucina pectinata/bacteria symbiosis. II. An electron paramagnetic resonance and optical spectral study of the ferric proteins. J. J. Biol. Chem. 265, 16054

Krell, T., Lecal, J., Busch, A., Silva-Jimenez, H., Guazzaroni, M., Ramos, J.L. (2010) Bacterial Sensor Kinases: Diversity in the Recognition of Environmental Signals. Annu. Rev. Microbiol. 64, 539-59.

Krissinel, E.B., Winn, M.D., Ballard, C.C., Ashton, A.W., Patel, P., Potterton, E.A., McNicholas, S.J., Cowtan, K.D., Emsley, P. (2004) The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. Acta Crystallogr D Biol Crystallogr. 60(Pt 12 Pt 1):2250-5

Kumaraswami M, Howe MM, Park HW Crystal Structure of the Mor Protein of Bacteriophage Mu, a Member of the Mor/C Family of Transcription Activators J. Biol. Chem., Apr 2004; 279: 16581 – 16590

Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss. 1989. Expression of sigma 54 (ntrA)-dependent genes is probably united by a common mechanism. Microbiol. Rev. 53:367-376.

Lowicka E, Beltowski J. (2007) Hydrogen sulfide (H_2S) = the third gas of interest for pharmacologists. Pharmacol Rep. 59, 4-24.

Lankas, F., Sponer, J., Langowski, J., Cheatham, T.E. III. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. Biophys. J. 85, 2872-83.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: Conventions and principles, J. Biomol. Struct. Dynam. 6, 655–667.

Lee, S.Y., De La Torre, A., Yan, D., Kustu, S., Nixon, B.T., Wemmer, D.E. (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. Genes Dev. 17, 2552-63.

Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., Bassler, B.L. (2004) The Small RNA Chaperone Hfq and Multiple Small RNAs Control Quorum Sensing in Vibrio harveyi and Vibrio cholerae. Cell, 118, 69-82.

Letain, T.E., Kane, S.R., Legler, T.C., Salazar, E.P., Agron, P.G., Beller, H.R. (2007) Development of a genetic system for the chemolithoautotrophic bacterium Thiobacillus denitrificans. Environ. Microbiol. 73, 3265-71.

Liebl, U., Bouzhir-Sima, L., Négrerie, M., Martin, J.L., Vos, M.H. (2002) Ultrafast ligand rebinding in the heme domain of the oxygen sensors FixL and Dos: general regulatory implications for heme-based sensors. Proc. Natl. Acad. Sci. 99, 12771–12776

Lisser, S., Margalit, H. (1993) Compilation of E. coli mRNA promoter sequences. Nuc. Ac. Res. 21, 1507-1516.

Loria. J.P. Rance, M., Palmer, A.G. III. (1999) A TROSY CPMG sequence for characterizing chemical exchange in large proteins. J Biomol NMR. 15, 151-5.
Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids res., 31, 5108–5121.
Lundberg, J.O., Weitzberg, E., Cole, J.A., Benjamin, N. (2004) Nitrate, bacteria and human health. Nat. Rev. Microbiol. 2, 593-602.
Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting Coled Coils from Protein Sequences Science 252:1162-1164.
Magasanik, B. (1993) The regulation of nitrogen utilization in enteric bacteria. J Cell Biochem. 51, 34-40.
Malhotra, A., Severinova, E., Darst S.A. (1996) Crystal structure of a s70 subunit fragment from E. coli RNA polymerase. Cell. 87, 127-136.
Marcia, M., Ermler, U., Peng, G., Michel, H. (2010) A new structure-based classification of sulfide:quinone oxidoreductases. Proteins. 78, 1073-83.
Marley J. Lu M. Bracken C. A method for efficient isotopic labelling of recombinant proteins. J. Biol. NMR, 20:71-75,2001}
Martinez-Argudo I, Little R, Shearer N, Johnson P, Dixon R. The NifL-NifA System: a multidomain transcriptional regulatory complex that integrates environmental signals. J Bacteriol. 2004 Feb;186(3):601-10.
McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. (2007) Phaser crystallographic software. J Appl. Crystallogr. 40, 658-674.
Merickel SK, Sanders ER, Luis-Vzquez-Ibar, Johnson RC. Subunit Exchange and the Role of Dimer Flexibility in DNA Binding by the Fis Protein. Biochemistry, 2002 41(18): 5788-5798.
Metzler, W.J., Wittekind, M., Goldfarb, V., Mueller, L., Farmer, II, B.T. (1996) Incorporation of 1H/13C/15N-{Ile, Leu, Val} into a Perdeuterated, 15N-Labeled Protein: Potential in Structure Determination of Large Proteins by NMR. J. Am. Chem. Soc. 118, 6800-6801.
Miyatake, H., Mukai, M., Park, S.-Y., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T., Shiro, Y. (2000) Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J. Mol. Biol. 301, 415-431.
Model P, Jovanovic G, Dworkin J. The Escherichia coli phage-shock-protein (psp) operon. Mol Microbiol. 1997 Apr; 24(2):255-61.
Montie, T.C., Doyle-Huntzinger, D., Craven, R.C., Holder, I.A. (1982) Loss of virulence associated with the absence of flagellum in an isogenic mutant of Pseudomonas aeruginosa in the burned-mouse model. Infect. Immun. 38, 1296-1298.
Moreno-Vivian, C., Cabello, P., Martinez-Luque, M., Blasco, R., Castillo, F. (1999) Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. J Bacteriol. 181, 6573-6584.
Morett E., Bork P. (1998) Evolution of new protein function: recombinational enhancer FIS originated by horizontal gene transfer from the transcriptional regulator NtrC. FEBS Letters 433, 108-112.

Mori, S., Abeygunawardana, C., O'Neil-Johnson, M., vanZijl, P.C.M. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. J. Magn. Reson. B 108, 94-98.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E., Dahlquist, F.W. "Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance." Methods in Enzymology, Academic Press, Volume 177, 44-73.

Niehus, E., Gressmann, H., Ye, F., Schlapbach, R., Dehio, M., Dehio, C., Stack, A., Meyer, T.F., Suerbaum, S. Josenhans, C. (2004) Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori. Mol. Microbiol. 52, 947-961.

Nienhaus, K., Nienhaus, G.U. (2005) Probing Heme Protein-Ligand Interactions by UV/Visible Absorption Spectroscopy. Methods in Mol. Biol. 305, 215-241.

Nietlispach, D. (2005) Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. J. Biomol. NMR. 31, 161-6.

Ninfa A.J., Reitzer, L.J., and Magasanik, B. (1987) Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers. Cell 50:1039-1046.

Ninnemann, I., Kock, C., Kahmann, R. (1992) The E. coli fis promoter is subject to stringent control and autoregulation. EMBO J. 11, 1075-1083.

North A, K. S. (1997). Mutant Forms of the Enhancer-Binding Protein NtrC can Activate Transcription from Solution. JMB 267: 17-36.

Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. PNAS 95, 11163-8.

Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. in Methods in Enzymology Vol. 276 (eds. Carter, C.W.J. & Sweet, R.M.) 307–325 (Academic Press, Boston, 1997).

Pabo, Carl O., Sauer, R.T. Protein-DNA Recognition. Ann. Rev. Biochem. 1984. 53:293-321.

Pacheco, A. J.R., Pena Cabriales, J.J., Maldonado Vega, M. (2008) Identification and characterization of sulfur-oxidizing bacteria in an artificial wetland that treats wastewater from a tannery. Int J Phytoremediation. 10, 359-70.

Paget, M.S., Helmann, J.D. (2003) The sigma70 family of sigma factors. Genome Biol. 4, 203.

Palmer III, A.G., Cavanagh, J., Wright, P.E., Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. J. Magn. Reson. 83, 151-170.

Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S., Johnson, R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts, J. Mol. Biol. 264, 675–695.

Pandini, A., and L. Bonati (2005) Conservation and specialization in PAS domain dynamics. Prot. Eng. Des. Sel. 18, 127-137.

Park, S., Meyer, M., Jones, A.D., Yennawar, H.P., Yennawar, N.H., Nixon, B.T. (2002) Two-component signaling in the AAA + ATPase DctD: binding Mg2+

and BeF3- selects between alternate dimeric states of the receiver domain. FASEB J. 16, 1964-1966.
Pelton J.G., Kustu S., Wemmer D.E. (1999) Solution Structure of the DNA-binding Domain of NtrC with Three Alanine Substitutions. J. Mol. Biol. 292, 1095-1110.
Pereira, L.E., Brachmachary, P., Hoover, T.R. (2006) Characterization of Helicobacter pylori σ54 promoter-binding activity. FEMS Microbiol Lett 259, 20-26.
Pervushin, K., Riek, R., Wider, G., Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. PNAS. 94, 12366-71.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605-12.
Piantadosi, C.A. (2008) Carbon monoxide, ractive oxygen signaling, and oxidative stress. Free Radic Biol Med. 45, 562-9.
Poon DK, Wither SG, McIntosh LP. Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi Xylanase Cex through NMR spectroscopic analysis. J Biol Chem. 2007 Jan 19;282(3):2091-100.
Porter, R. I., North, A.K., Kustu, S. (1995) Mechanism of transcriptional activation by NtrC, p. 147-158. In J. A. Hoch and T. J. Silhavy (ed.), Two-component signal transduction. American Society for Microbiology, Washington, DC.
Porter, S.C., North, A.K., Kustu S. (1993) Oligomerization of NTRC at the glnA enhancer is required for transcriptional activation. Genes Dev. 7, 2258-73.
Raetz, C.R.H., Whitfield, Chris. (2002) Lipopolysaccharide Endotoxins. Annu. Rev. Biochem. 71, 635-700.
Ramos, J.L., Marqués, S., Timmis, K.N. (1997) Transcriptional control of the Pseudomonas TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annu. Rev. Microbiol. 51, 341-73.
Rappas M, Bose D, Zhang X. Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. Curr Opin Struct Biol. 2007 Feb;17(1):110-6.
Rappas, M., Schumacher, J., Beuron, F., Niwa, H., Bordes, P., Wigneshweraraj, S., Keetch, C.A., Robinson, C.V., Buck, M., Zhang, X. (2005) Structural insights into the activity of enhancer-binding proteins. Science. 307, 1972-5.
Rappas, M., Schumacher, J., Niwa, H., Buck, M., Zhang, X. (2006) Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. J Mol Biol. 357, 481-92.
Ray, P., Smith, K.J., Parslow, R.A., Dixon, R., Hyde, E.I. (2002) Secondary structure and DNA binding by the C-terminal domain of the transcriptional activator NifA from Klebsiella pneumoniae. Nuc Ac Res. 30, 3972-80.

Reitzer, L.J., Schneider, B.L. (2001) Metabolic Context and Possible Physiological Themes of sigma 54-Dependent Genes in Escherichia coli. Microbiology and Molecular Biology Reviews, 65, 422-444.

Reizer L.J., M. B., Magasanik B (1989). Activation of glnA transcription by nitrogen regulator I (NRI)-phosphate in Escherichia coli: evidence for a long-range physical interaction between NRI-phosphate and RNA polymerase. Journal of Bacteriology 171(10): 5512-22.

Révet B, Brahms S, Brahms G. (1995) Binding of the transcription activator NRI (NTRC) to a supercoiled DNA segment imitates association with the natural enhancer: an electron microscopic investigation. Proc Natl Acad Sci 92, 7535-9.

Rice, P., Longden, I., Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite Trends in Genetics 16, 276-277.

Rogers, K.R., Lukat-Rodgers, G.S. (2005) Insights into heme-based O_2 sensing from structure–function relationships in the FixL proteins. J Inorg Biochem. 99, 963-977.

Sakurai, H., Ogawa, T., Shiga, M., Inoue, K. (2010) Inorganic sulfur oxidizing system in green sulfur bacteria. Photosynth Res. 104, 163-676.

Sallai, L., Tucker, P.A. (2005) Crystal structure of the central and C-terminal domain of the sigma54-activator ZraR. J. Struct. Biol, 151, 160-170.

Santero, E., Hoover, T.R., North, A.K., Berger, D.K., Porter, S.C., Kustu, S. (1992) Role of integration host factor in stimulating transcription from the sigma54-dependent nifH promoter. J Mol Bio. 227, 602-20.

Scholl D, Nixon T. Cooperative Binding of DctD to the dctA Upstream Activation Sequence of Rhizobium meliloti Is Enhanced in a Constitutively Active Truncated Mutant. JBC (1996) 271 (42):26435-26442.

Schulte-Herbrueggen, T., Sorensen, O.W. (2000) Clean TROSY: compensation for relaxation-induced artifacts. J. Magn. Reson. 144, 123-128.

Schulz A, Langowski J, Rippe K. The Effect of the DNA Conformation on the Rate of NtrC activated Transcription of Escherichia coli RNA Polymerase*sigma54 Holoenzyme. Journal of Molecular Biology. 300(4) 21 July 2000: 709-725.

Schumacher J, Joly N, rappas M, Zhang X, Buck M. Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation. J struct Biol. 2006 Oct; 156(1):190-9.

Sean P, Nguyen JH, Semler BL. The linker domain of poly(rC) binding protein 2 is a major determinant in poliovirus cap-independent translation. Virology. 2008 Sep 1;378(2):243-53.

Setchell, W. A. (1903) The upper temperature limits of life. Science 17, 934–937.

Shahak, Y., B. Arieli, E. Padan, and G. Hauska. (1992) Sulfide quinone reductase (SQR) activity in Chlorobium. FEBS Lett. 299:127-130

Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H., Kay, L.E. (1996) Assignment of 15N, 13Cα, 13Cβ, and HN Resonances in an 15N,13C,2H Labeled 64 kDa Trp Repressor-Operator Complex Using Triple-Resonance NMR Spectroscopy and 2H-Decoupling. J Am. Chem. Soc. 118, 6570-6579.

Sheely, M.L. Ind. Eng. Chem., 1932, 24 (9), pp 1060–1064.

Spirin, A.S., Baranov, V.I., Ryabova, L.A., Orodov, S.Y., Alakhov, Y.B. (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. Science. 242: 1162-1164.
Staunton, D., Schlinkert, R., Zanetti, G., Colebrook, S.A., Campbell, I.D. (2006) Cell-free expression and selective isotope labeling in protein NMR.
Stella, S. Cascio, D. Johnson, R.C. (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes Dev. 24, 814-826.
Stetter, K.O. (2006) Hyperthermophiles in the history of life. Philos Trans R Soc Lond B Biol Sci. 361, 1837-1843.
Studholme, D.J., Buck, M. (2000) The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. FEMS microbial. let. 186, 1.
Studholme, D.J., Dixon, R. (2003) Domain Architectures of s54-Dependent Transcriptional Activators. J Bacteriol. 185, 1757-1767.
Studholme, D.J., Wigneshwereraraj, S.R., Gallegos, M.T., Buck, M. (2000) Functionality of purified sigma(N) (sigma(54)) and a NifA-like protein from the hyperthermophile *Aquifex aeolicus*. J. Bacteriol. 182, 1616-23.
Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expression Purif.41, 207-234.
Su W, P. S., Kustu S, Echols H (1990). DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. PNAS 87(14): 5504-8.
Swint-Kruse L, Matthews KS. Allostery in the LacI/GalR family: variations on a theme. Curr Opin Microbiol. 2009 Apr;12(2):129-37.
Thompson, J.F., Landy, A. (1988) Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes, Nucleic Acids Res. 16, 9687–9705.
Truong, D.H., Eghbal, M.A., Hindmarsh, W., Roth, S.H., O'Brien, P.J. (2006) Molecular mechanisms of hydrogen sulfide toxicity. Drug Metab Rev 38, 733-44.
Tucker, N.P., Ghosh, T., Bush, M., Zhang, X., Dixon, R. (2010) Essential roles of three enhancer sites in sigma54-dependent transcription by the nitric oxide sensing regulatory protein NorR. Nuc. Ac. Res. 38, 1182-94.
Tucker, P.A., Sallai, L. (2007) The AAA+ superfamily – a myriad of motions. Curr Opin Struct Biol. 17, 641-52.
Tugarinov, V., Muhandiram, R., Ayed, A., Kay, L.E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. JACS. 124, 10025-35.
Tusnády, G.E., and I. Simon (1998) Principles Governing Amino Acid Composition of Integral Membrane Proteins: Applications to Topology Prediction." J. Mol. Biol. 283, 489-506.
Tusnády, G.E., and I. Simon (2001) The HMMTOP transmembrane topology prediction server" Bioinformatics 17, 849-850.
Ussery, D., Larsen, T.S., Wilkes, K.T., Friis, C., Worning, P., Krogh, A., Brunak, S. (2001) Genome organization and chromatic structure in Escherichia coli. Biochimie 83, 201-212.

Valverde, C., Haas, D. (2008) Small RNAs Controlled by Two-Component Systems. Adv. Exp. Med. Biol. 631, 54-79.
van Dijk, M., Bonvin, A.M.J.J. (2009) "3D-DART: a DNA structure modelling server", Nucl. Acids Res., 37 (Web Server Issue):W235-W239
van Dijk, M., van Dijk, A. D., Hsu, V., Boelens, R., Bonvin, A. M. (2006) Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Res, 34, 3317-25.
Vasilaki E, Siderakis M, Papakosta P, Skourti-Stathaki K, Mavridou S, Kardassis D. Novel regulation of Smad3 oligomerization and DNA binding by its linker domain. Biochemistry. 2009 Sep 8;48(35):8366-78.
Vuister, G.W., Kim, S.J., Wu, C., Bax, A. (1994) 2D and 3D NMR-study of phenylalanine residues in proteins by reverse isotopic labeling. J Am Chem Soc. 116, 9206-9210.
Wedel A, Popham, D., Droge, P., Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. Science 248(4954): 486-90.
Wedel, A., Kustu, S. (1995) The bacterial enhancer-binding protein NTRC is a molecular machine: ATP hydrolysis is coupled to transcriptional activation. Genes Dev. 9, 2042-52.
Wedel, A., Weiss, D.S., Popham, D., Droege, P., Kustu, S. (1990) A bacterial enhancer functions to tether a transcriptional activator near a promoter. Science. 248, 486-90.
Wei, L., Stevenson, C.E.M., Burton, N., Jakimowicz, P., Paget M.S.B., Buttner, M.J., Lawson, D.M. Kleanthous, C. (2002) Identification and structure of the anti-sigma factor-binding domain of the disulphide-stress regulated sigma factor factor sigmaR from Streptomyces coelicolor. J Mol Biol. 323, 225-236.
Weiss, V., Günter Kramer, Thomas Dünnebier, and Annette Flotho (2002) Mechanism of Regulation of the Bifunctional Histidine Kinase NtrB in Escherichia coli. J Mol Microbiol Biotech. 4, 229-33.
Wider, G., Wuthrich, K. (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. Curr. Op. Struct. Biol. 9, 594-601.
Wigneshweraraj, S., Bose, D., Burrows, P.C., Joly, N., Schumacher, J., Rappas, M., Pape, T., Zhang, X., Stockley, P., Severinov, K., Buck, M. (2008) Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. Mol Microbiol. 68, 538-46.
Wong, C., Tintut, Y., Gralla, J. (1994) The Domain Structure of Sigm54 as Determined by Analysis of a Set of Deletion Mutants. J. Mol. Biol. 236, 81-90.
Wosten, M.M. (1998) Eubacterial sigma-factors. FEMS Microbiol Rev. 22, 127-50.
Yamada, S., Sugimoto, H., Kobayashi, M., Ohno, A., Nakamura, H., Shiro, Y. (2009) Structure of PAS-linked histidine kinase and the response regulator complex. Structure. 17, 1333-1344
Yuan, H.S., Finkel, S.E., Feng, J-A., Johnson, R.C., Dickerson, R.E. (1991) The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. Proc. Natl. Acad. Sci. 88, 9558-9562

Yuan, H.S., Wang, S.S., Yang, W.Z., Finkel, S.E., Johnson, R.C. (1994) The Structure of FIS Mutant PRO61Ala illustrates that the kink within the long alpha-helix is not due to the presence of the proline residue. J. Biol. Chem. 269, 28947-28954.

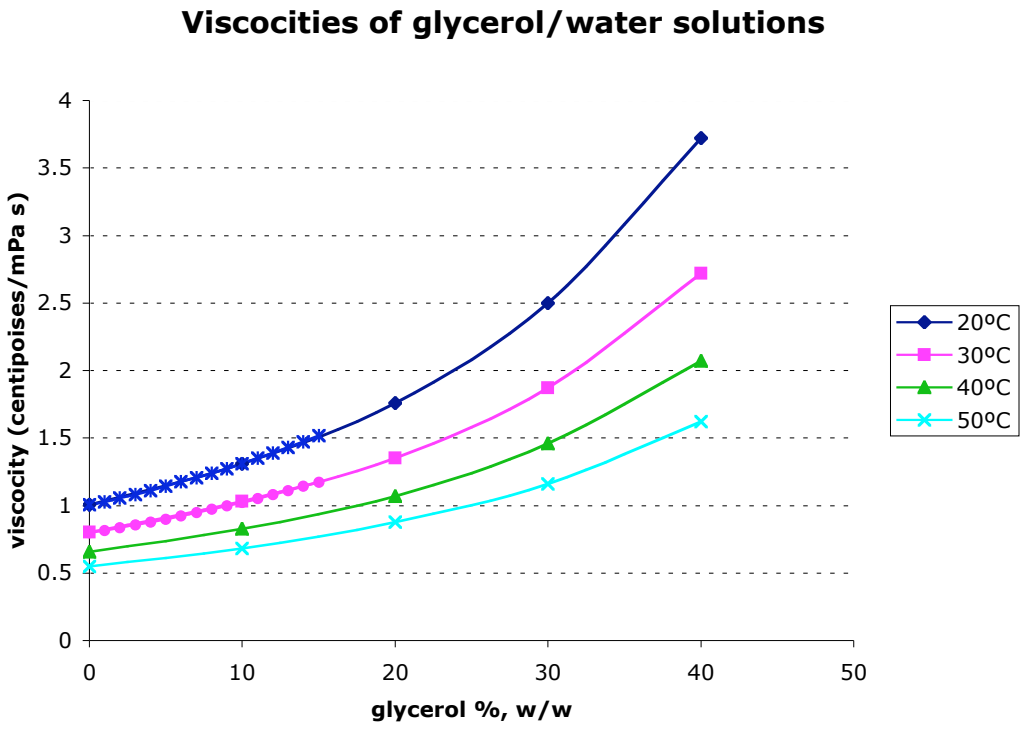
Zhang X, Chaney M, Wigneshweraraj SR, Schumacher J, Bordes P, Cannon W, Buck M. Mechanochemical ATPases and transcriptional activation. Mol Microbiol. 2002 Aug;45(4):895-903.

Zimmer D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter B.J., Bender, R.A., Kustu, S. (2000) Nitrogen regulatory protein C-controlled genes of Escherichia coli: scavenging as a defense against nitrogen limitation. Proc Natl Acad Sci 97, 14044-5.

Appendix A. *Tabulation of viscocity of glycerol solutions at varying temperatures. Black values are from Dorsey, 1940. Blue values are from Sheely, 1932.*

% glycerol w/w in H_2O solution	*viscosity (centipoises/mPa•s)*			
	20ºC	30ºC	40ºC	50ºC
0	1.005/1.005	0.800/0.8007	0.656	0.5494
1	1.029	0.912	---	---
2	1.055	0.935	---	---
3	1.083	0.959	---	---
4	1.112	0.984	---	---
5	1.143	1.010	---	---
6	1.175	1.037	---	---
7	1.207	1.064	---	---
8	1.239	1.092	---	---
9	1.274	1.121	---	---
10	1.311/1.31	1.024/1.03	.826	0.69
20	1.76	1.35	1.07	0.879
30	2.5	1.87	1.46	1.16
40	3.72	2.72	2.07	1.62

Scatter plot of viscocities of data in Table 3.



Appendix B. Communication from Professor Valley Stewart, UC Davis, on the function of genes in the nirB and fhp gene clusters.

These genes appear to be involved in nitrate assimilation. I strongly doubt that they are involved in nitrate respiration.

nirB: encodes NADH-dependent nitrite reductase (nitrite reduced to ammonium). This enzyme is conserved in fungi as well as bacteria, and has been studied in both. In many species, the enzyme is composed of two subunits (NirB and NirD, as in E. coli), but in others the gene is fused to make a single polypeptide (as in K. pneumoniae). I do not know of any case where NirBD enzyme is involved in nitrate respiration (energy conservation). In E. coli and some other enterobacteria it plays a role in NAD+ regeneration and/or in nitrite detoxification; however, it almost always is associated with nitrate assimilation.

cobA: encodes an enzyme that forms a branchpoint off of the mainline heme biosynthetic pathway. This branchpoint is used to manufacture the porphryin moiety of vitamin B12 (cobalamin), and so this gene is often annotated as "cobA." However, the same branchpoint makes siroheme, which is the cofactor for NirBD enzyme. In E. coli, this enzyme is encoded by the cysG gene, because siroheme is also the cofactor for sulfite reductase.

trpD2: has no obvious relationship to nitrate metabolism. However, the TrpG (amino-terminal) portion of the TrpD polypeptide (ancestral fusion of TrpG and TrpD) comprises the glutamine amidotransferase step of anthranilate synthesis. Thus, its inclusion in this operon may be related to overall glutamine (nitrogen) homeostasis.

fhp: encodes only the globin portion of flavohemoprotein, which is an FAD/NAD enzyme involved in nitric oxide detoxification. NO is a byproduct of nitrite reduction to ammonium, so it is not unusual for assimilatory nitrate reduction operons to encode nitric oxide-metabolizing enzymes. Pure speculation would be that this globin fragment is involved in sequestering NO, or that other components of a detoxifying enzyme complex are encoded elsewhere.

cynS: is involved in nitrogen compound metabolism, but has no obvious relationship to nitrate metabolism.

glnB: you know about.

nasA: encodes a major facilitator subfamily protein (12-transmembrane helix protein) that transports nitrate across the cytoplasmic membrane. This functions as a nitrate-proton symport for nitrate assimilation. An alternate annotation is narK. LacY is the paradigm for this widespread family of proteins.

narB: encodes ferredoxin-nitrate reductase, which is involved in nitrate assimilation in cyanobacteria.

So, this operon has everything the organism needs to assimilate nitrate into ammonium. None of these enzymes would function for nitrate respiration ("dissimilation.").

-- Professor Valley Stewart, UC Davis, 1/25/2010