Mechanisms of σ^{54} bacterial transcription activation

by

Alexander Rigel Siegel

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

in the Graduate Division
of the University of California, Berkeley

Committee in charge:
Professor David E. Wemmer, Chair
Professor John Kuriyan
Professor Andreas Martin
Professor Donald C. Rio

Summer 2016

Abstract

Mechanisms of σ^{54} bacterial transcription activation

by

Alexander Rigel Siegel

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor David E. Wemmer, Chair

This dissertation addresses the mechanism of σ^{54} activation by the AAA+ ATPase transcriptional activators. The first chapter provides a general introduction to σ^{54}-mediated bacterial transcription initiation by outlining the existing structural and biochemical data on σ^{54} and its activators. The majority of our structural information comes from high resolution structures of individual σ^{54} and transcriptional activator domains, with the rest coming from low resolution structures that determine relative positions of the domains. The structural and biochemical properties of these domains, viewed in the context of the mechanisms of related motor proteins, led to our hypothesis of σ^{54} activation. In this dissertation, I propose that the transcriptional activator's hexameric ATPase domain uses conserved loops, known to contact σ^{54}, to pull on and thread the σ^{54} N-terminal activator interacting domain through its central pore. In this model, through processive rounds of ATP hydrolysis, the transcriptional activator applies enough force to generate a conformational change in the σ^{54}-RNA polymerase holoenzyme that allows it to melt DNA thus initiating transcription. The primary goal of my work has been the study of the activation mechanism using a variety of techniques, including traditional NMR-based structure determination, *in vivo* and *in vitro* biochemistry, and single molecule optical tweezers experiments.

The second chapter outlines the central work of this dissertation, the structural characterization of the region of σ^{54} responsible for contacting the activator and necessary for initiating transcription. We show that the activator interacting domain

(AID) is intrinsically disordered, but becomes ordered when bound to the transcriptional activator, NtrC1, in its ATP state. In particular, we show that two predicted helices in the sequence are sufficient for activator binding with native-like affinity, and that the first helix in particular represents the primary region of contact. We applied TROSY-based NMR techniques to the structure determination of this high molecular weight complex, but most signals were broad due to dynamics or disorder, preventing a high resolution structure determination of the AID bound to the activator. This indicates that the AID does not bind the activator in a single conformation, but rather in multiple conformations each experiencing a different chemical environment. This may reflect an activation mechanism involving dynamic changes to the σ^{54}-activator interaction site.

The third chapter takes an *in vivo* biochemical approach to study the σ^{54} activation mechanism by characterizing the functionality of σ^{54} with insertions and deletions near the activator interacting domain. These insertions distinguish between two competing mechanisms: whether the σ^{54} AID only binds the surface of the activator or, as we propose, is threaded through its central pore. We find that a flexible linker region between the sites of activator and RNA polymerase binding is essential for fully functional σ^{54}. While the length of the linker does not matter, the insertion of a small, stably folded domain between these two interaction sites is detrimental to σ^{54} activity *in vivo*. This is consistent with the threading of the AID and linker by the transcriptional activator, which would not be inhibited by the addition of extra unstructured residues but would be stalled by the addition of a folded domain in the linker.

The fourth chapter outlines single molecule optical tweezers experiments to characterize the effect of force on a domain of σ^{54}. In particular, we examine the possibility that the σ^{54} core binding domain (CBD), responsible for contacting core RNA polymerase, acts as a conformational fracture point that undergoes rearrangements when force is applied. The NMR structure of the CBD revealed two folded subdomains with a hydrophobic interface that could plausibly be disrupted by the force of the activator threading the N-terminus. By applying force with optical tweezers to either end of the CBD alone, we observed a separate unfolding event likely corresponding to unfolding of its seventh helix, but there was no evidence that the less stable of the two CBD subdomains unfolds before the other. Future experiments to conclusively test the force-dependent activation of σ^{54} may require reconstituting and pulling on the full RNA polymerase holoenzyme complex, including core RNAP, promoter DNA, and full length σ^{54}.

The fifth chapter steps back from the study of the σ^{54} activation by the transcriptional activators to examine the regulation of the transcriptional activators themselves. Existing evidence shows that phosphorylation of the regulatory receiver domain of the transcriptional activator NtrC changes the population of its active and inactive states, with most states resembling the active conformation after phosphorylation. The activated receiver domain promotes the oligomerization of the ATPase domain, which in turn can activate σ^{54}. In this chapter, I studied an NtrC receiver domain from the piezophilic bacteria, *Shewanella violacea*, which turns on σ^{54}-dependent gene expression in response to high pressure. We hypothesized that pressure alone might drive the conformational change normally associated with phosphorylation, thereby activating *S.v.* NtrC independent of its normal, two-component signaling pathway. Using high pressure NMR spectroscopy, we showed that increased pressure alone does not increase the population of the *S.v.* NtrC receiver domain's active state. Future work must consider the pressure sensing behavior of the other components in the NtrC signaling pathway.

To Ronnie, Peter, and Heather, whose love supported me through the many trials of research.

Acknowledgments

I'd like to thank Heather Lee and my parents, Ronnie and Peter Siegel, for the love and support they've given me. I'd like to thank my many mentors, in particular my advisor, David Wemmer, for his patience and persistence helping me overcome the many pitfalls of my projects. I have a deep admiration for his breadth of knowledge and the rigorous way he goes about solving the most challenging scientific problems. I'm also grateful to Jeff Pelton for his excellent instructions on setting up and running NMR experiments and to Kwang Seo Kim for teaching me many fundamental biochemistry and molecular biology techniques. I'd like to thank my thesis committee, Andreas Martin, John Kuriyan and Donald Rio, for the helpful feedback.

I'd like to thank the many contributors to the projects discussed in this dissertation. In Chapter 2, Seh-hyeon Jin and Zhijuan Gao contributed to the studies of the *E. coli* σ^{54} core binding domain, Cristhian Canari contributed to the studies of σ^{54}-RNAP binding, and Kwang Seo Kim contributed to the fluorescence anisotropy measurements of full length σ^{54}. In Chapter 3, Kwang Seo Kim contributed to the cloning and knock outs of the rpoN gene. In Chapter 4, Laura Rosen and Matthew Draelos performed the circular dichroism and hydrogen exchange experiments on the core binding domain, Wenshu Wang, Katie Amberg-Johnson and Jeaneun Park contributed to the setup of the molecular tweezers experiments, and Jessie Dill helped collect and analyze the tweezer's data. In Chapter 5, Werner Kremer collected the NMR data at high pressure.

I'd like to thank my many labmates over the years. My friend and colleague throughout most of grad school, Zhijuan Gao, who was always willing to discuss ideas and share reagents. My undergraduate researcher Katie Amberg-Johnson who contributed three years of excellent work and a cheerful attitude towards science that brightened the whole lab. Joe Batchelor, who welcomed me to the Wemmer Lab and introduced me to many of the techniques I'd use in grad school. All of my other labmates over the years who made the Wemmer Lab a pleasant and collaborative place to work, including Aaron Phillips, Natasha Keith Vidangos, Monica Smith, Cathleen Zeymer, Rahul Das, Sonya Lorenz, Wenshu Wang, Aleks Kijac, Hagit Sorek, Beibei Huang, Seh-hyeon Jin, Matthew Draelos, Kevin O'Fee, George Shan, David Casas-Mao, Keunhong Jeong, and Jeaneun Park. I thank the members of the neighboring Marqusee, Kuriyan, and Martin labs for the friendly and collaborative environment in Stanley Hall. I thank the Biophysics Graduate Group, which has been a great community to share ideas, and in particular Kate Chase for looking after the needs of all the Biophysics students.

Table of Contents

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

1.1 Bacterial Transcription

All essential cellular processes, from metabolism to replication, depend on the cell's ability to produce the correct proteins in the correct amounts at the correct times and send them to the correct places. The majority of these enzymes are proteins, which fold in a sequence-specific manner and are capable of catalyzing the huge number of different biochemical reactions necessary for cell survival and proliferation. The information needed to produce the exact sequence of each protein is stored in the DNA, then transcribed to mRNA, which is in turn translated to proteins. However, to create the proper balance of biochemical reactions it is not enough to just make the correct proteins, the cell also needs to modulate their concentrations during its growth cycle and in response to the environment. In bacteria, while the regulation of gene expression can occur during translation and post-translationally, the most dynamic point of control of gene expression occurs at the level of transcription.

Transcription is performed by core RNA polymerase (RNAP), a 400 kDa protein complex which consists of five core subunits: two α subunits, β, β', and ω. While core RNA polymerase is capable of elongating an mRNA strand from single stranded template DNA, it is not able to bind or melt the double stranded promoter sequences along the DNA upstream of the gene's start site (Buck et al. 2000). In order to recognize and melt these promoter sequences, RNAP must bind to a sixth, modular subunit called sigma factor to form the RNA polymerase holoenzyme (Burgess et al. 1969). There are often many different sigma factors in the cell, and each one confers the holoenzyme with different specificity for promoter regions and helps position it in front of the correct genes before transcription (Losick and Pero 1981). The sigma factors also assist the holoenzyme in melting the DNA at the promoter site to grant the core RNAP access to a single strand of DNA in front of the gene, which it uses as a template to transcribe the complementary RNA.

The sigma factors are an important transcriptional regulatory component of gene expression. The concentration of each of the different sigma factors (seven in *E. coli*) is also controlled by regulating their transcription and translation, as well as

post-translational processing and targeting them for degradation (Paget and Helmann 2003). The sigma factors can also be inactivated by expressing anti-sigma factors, which bind them and prevent their association with core RNAP (Campbell, Westblade, and Darst 2008). Repressors or enhancers can also bind at or near the gene to either block or recruit certain sigma factors from accessing the promoter sequences (Colland et al. 1998).

The essential role of sigma factors in DNA recognition and melting makes them an important class of proteins to study. Though eukaryotic transcription uses a different cast of proteins, understanding the fundamental biological mechanisms performed by the sigma factor and RNAP, like the process of melting DNA and the activation mechanisms of transcriptional activators, will provide structural insights that apply to transcription in higher organisms as well.

1.2 Comparison of transcription initiation by σ^{70} and σ^{54}

Bacterial sigma factors fall into one of two structurally distinct classes named after the molecular weight of their *E. coli* homologs: σ^{70} and σ^{54}. Though members of both classes direct the RNAP holoenzyme to the promoter sequences in front of genes and melt the DNA, they have different amino acid sequences, domain architectures, and mechanisms of initiating transcription. Members of the more numerous σ^{70} class control gene expression of most proteins needed to perform typical cellular functions, while members of the σ^{54} class more often control gene expression of proteins required to deal with sudden and unexpected environmental changes (Kazmierczak, Wiedmann, and Boor 2005).

Sigma factors complex with RNAP to form the holoenzyme which binds two DNA sites in the promoter region upstream of genes. Upon assembling at the promoter sites, σ^{70}-RNAP holoenzyme is immediately active and able to melt DNA and initiate transcription (Browning and Busby 2004)(Ghosh, Bose, and Zhang 2010). σ^{54} requires an additional activation step, where a member of a class of AAA+ ATPase transcriptional activators assembles further upstream and activates σ^{54} through ATP hydrolysis (Wedel and Kustu 1995). This activation drives a conformational change that allows the σ^{54}-RNAP to melt the DNA and initiate transcription. Once the DNA has been opened the sigma factors can dissociate, while the core RNAP continues to

elongate RNA using the single stranded DNA template containing the gene (Korzheva et al. 2000)(Gnatt et al. 2001)(Tahirov et al. 2002) (Figure 1.1). The additional activation requirement gives σ^{54} tighter control over gene expression, allowing for both a rapid response and reduced levels of background expression.

σ^{70} activity is primarily regulated by its own concentration and by the concentration of anti-sigma factors that bind and inactivate it (Campbell, Westblade, and Darst 2008). Consequently, σ^{70} gene expression is difficult to turn off entirely or to turn on suddenly, because modulating the expression of new σ^{70} or anti-sigma factors and targeting them for degradation are relatively slow processes. σ^{54}, on the other hand, is primarily regulated by the activity of its transcriptional activators, which are often already present in the cell but remain inactive until they receive a signal. σ^{54}-dependent gene expression can be turned on quickly by a signaling pathway without waiting for the expression of new regulatory proteins. Similarly, the signaling pathways can shut off σ^{54}-dependent gene expression quickly by deactivating the transcriptional activators without requiring a delay to target regulatory proteins for degradation. This difference in activation distinguishes the roles of σ^{70} and σ^{54} dependent gene expression in the cell: σ^{70} controls the expression of housekeeping genes which can appropriately have a steady level of expression (Lonetto, Gribskov, and Gross 1992), while σ^{54} controls the expression of genes required for mounting an immediate response to environmental changes (Kazmierczak, Wiedmann, and Boor 2005).

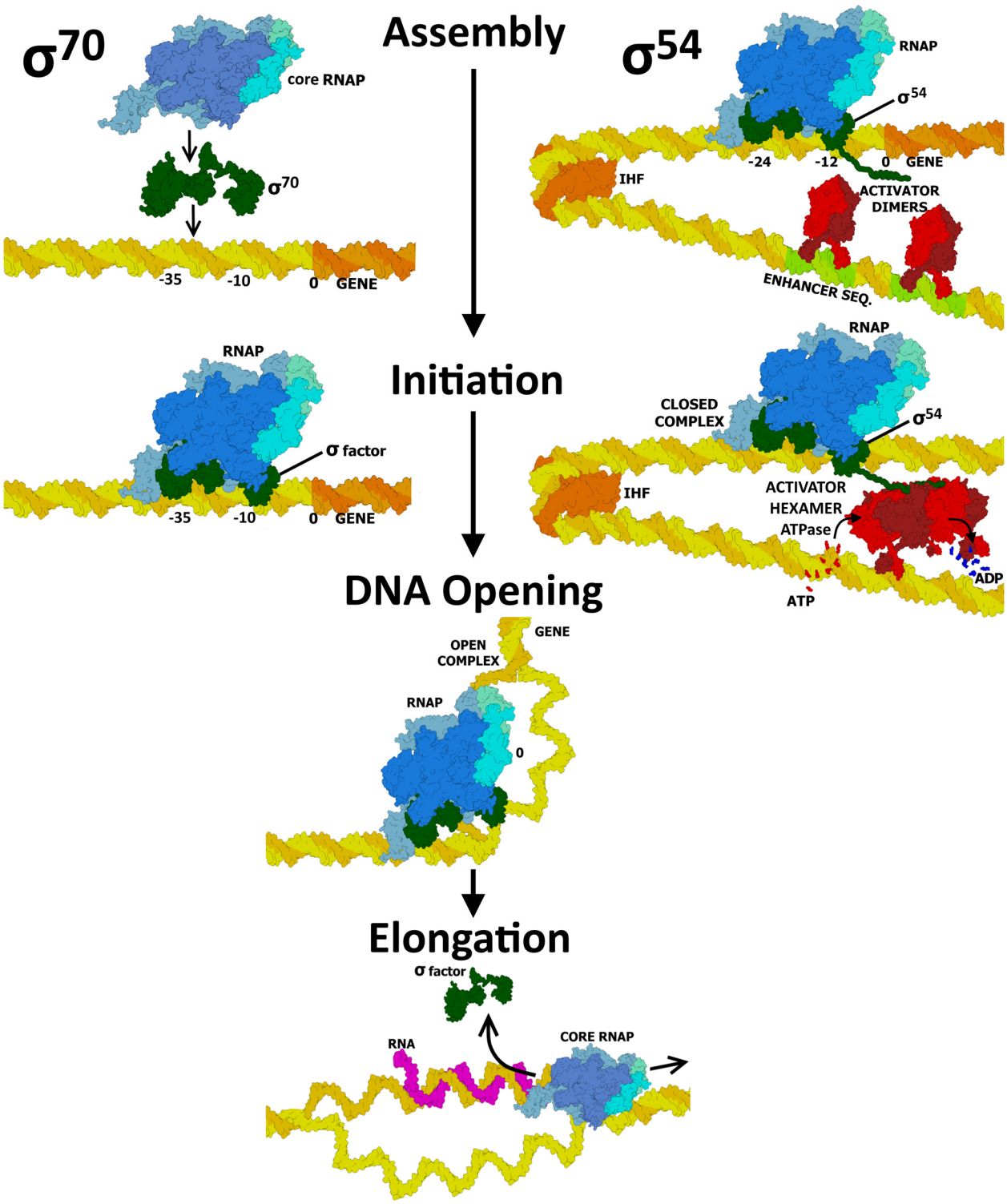


Figure 1.1 Diagram of transcription initiation mediated by the two classes of bacterial sigma factors. (1) Assembly: sigma factor and RNA polymerase assemble upstream of the start site. (2) Initiation: σ^{70} is immediately able to initiate transcription, while σ^{54} requires an activation event from one of the active, hexameric transcriptional activators. (3) DNA opening: the RNA polymerase holoenzyme melts DNA. (4) Elongation: the sigma factor falls off and core RNA polymerase continue to transcribe RNA off the single stranded DNA.

1.3 Structural overview of the σ^{70}-RNA polymerase holoenzyme

Structures of the σ^{70}-RNA polymerase holoenzyme have revealed details of how the holoenzyme forms the open complex and initiates transcription. σ^{70} has four individually folded domains separated by flexible linkers, referred to as σ_1, σ_2, σ_3, and σ_4 from the N- to the C-terminus (Murakami and Darst 2003). With the exception of the poorly conserved σ_1 domain, structures of the other σ^{70} domains are very similar in different species (Murakami et al. 2002)(Vassylyev et al. 2002)(Malhotra, Severinova, and Darst 1996). These structures reveal how σ^{70} binds to two DNA sites upstream of the transcription initiation site: TTGACA at the -35 promoter element and TATAAT at the -10 promoter element (Lisser and Margalit 1993). The σ^{70} σ_2 and σ_3 domains are responsible for -10 promoter recognition and a subregion of σ_2 called $\sigma_{2.3}$ melts it to form the open complex (Juang and Helmann 1994)(Barne et al. 1997). The σ_4 domain recognizes the -35 promoter element to position the holoenzyme properly at the promoter (Campbell et al. 2002)(Vassylyev et al. 2002). The poorly conserved σ_1 domain is negatively charged and may have a role in blocking the entry of DNA into the interior of RNAP until the DNA has been melted (Murakami et al. 2002)(Paget and Helmann 2003). In addition to their roles in DNA recognition, regions σ_2, σ_3, and σ_4, all provide significant contacts to core RNAP (Figure 1.2).

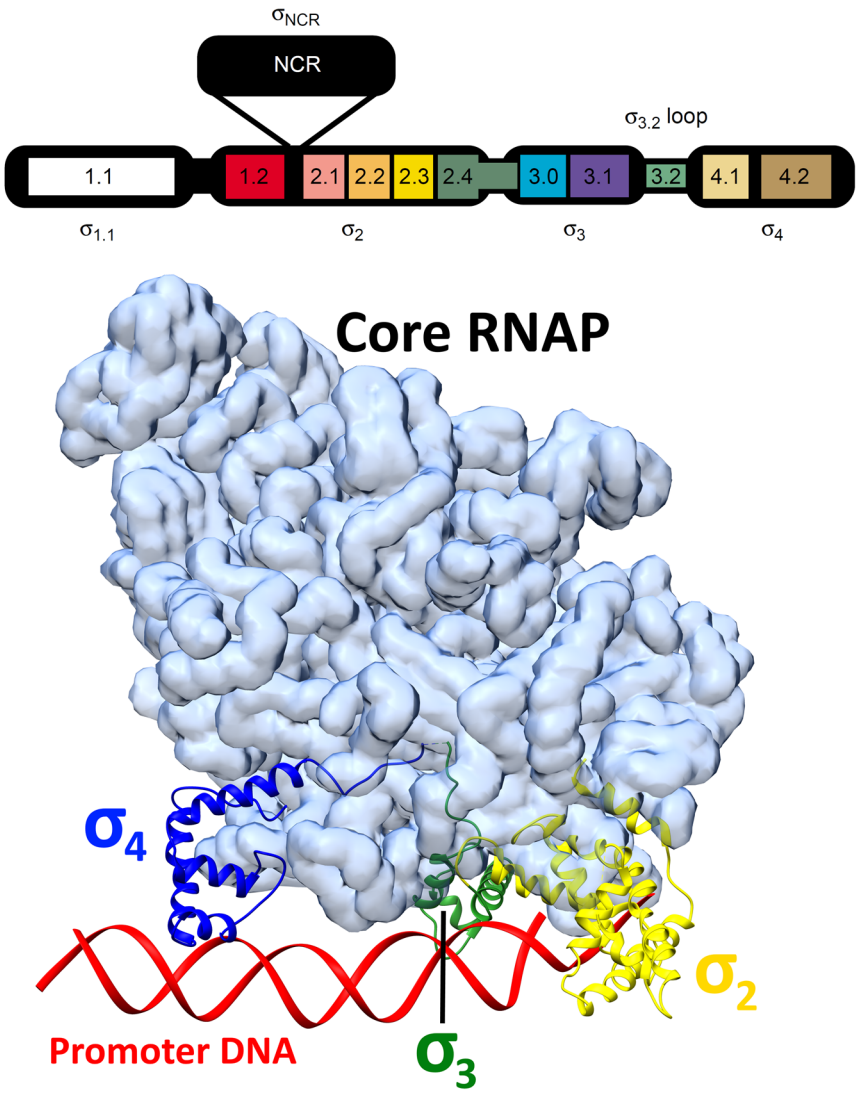


Figure 1.2 Domain architecture and structure of the σ^{70}-RNAP holoenzyme bound to DNA. Core RNAP is shown as a surface model (transparent light blue). The N-terminal $\sigma_{1.1}$ domain is not present. The domains σ_2 (yellow), σ_3 (green) and σ_4 (blue) are shown as ribbons bound to the core RNAP and to the -35 and -10 promoter elements along the DNA (red). PDB ID: 1L9Z (Murakami et al. 2002). Adapted from (Murakami and Darst 2003).

The structures of the σ^{70}-holoenzyme bound to DNA suggest a mechanism for σ^{70}-mediated transcription initiation (Murakami and Darst 2003) (Figure 1.3). Core RNAP binds to σ^{70}, which in turn binds to the -35 and -10 promoter elements upstream of the gene forming a closed complex with double stranded DNA. This positions a number of aromatic residues in the $\sigma_{2.3}$ region near the -10 promoter element. A natural breathing of the double-stranded, AT-rich DNA allows the aromatic residues in $\sigma_{2.3}$ to interact with nucleotide bases along a single strand of the -10 promoter element. This stabilizes the initial melting of the DNA, which then expands from the -11 position past the transcription start site to the +4 position to form the open complex. At this point

the single-stranded DNA is pushed deeper into the holoenzyme active site where RNAP begins to transcribe complementary RNA off of the single-stranded DNA template (Murakami et al. 2002). Once the RNA reaches 12 nucleotides in length, it destabilizes interactions between σ_4 and RNAP, leading to release of σ^{70}. Core RNAP forms the elongation complex, releasing the promoter and continuing to elongate RNA (Liu and Martin 2002)(Tahirov et al. 2002).

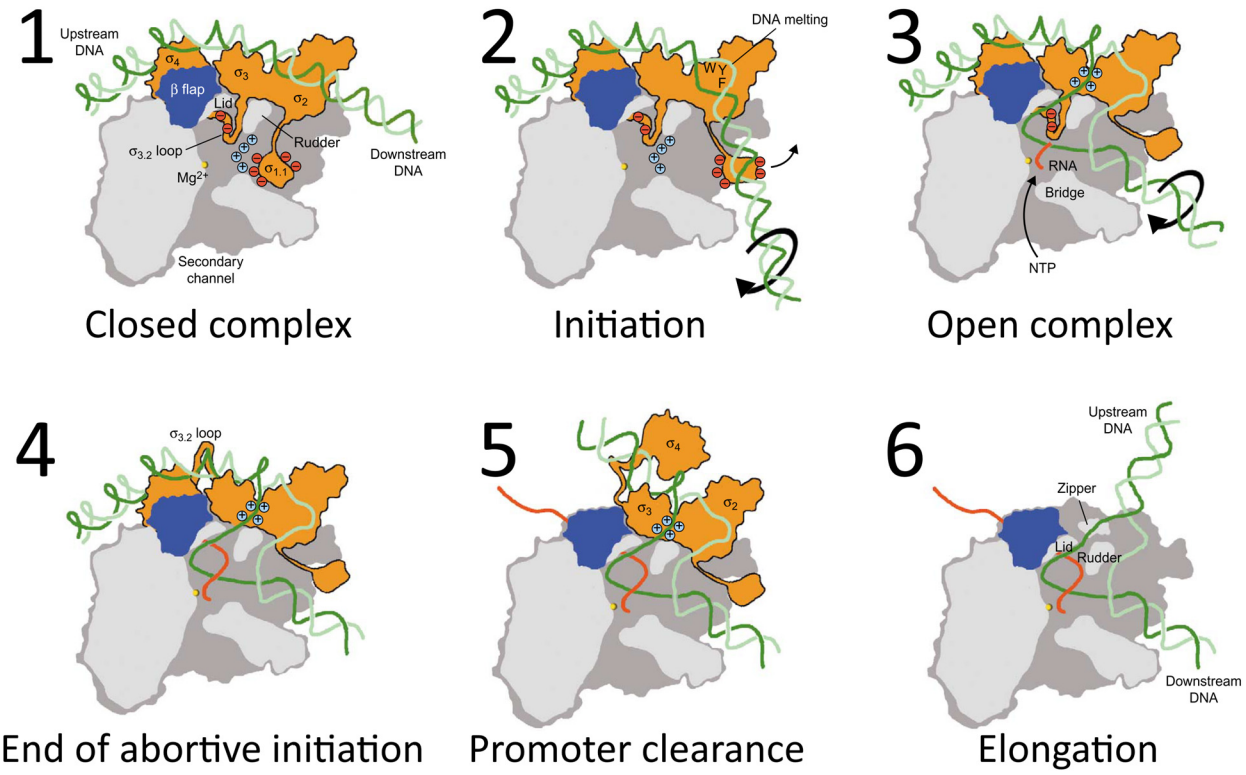


Figure 1.3 Cartoon diagram showing the steps in transcription initiation by the σ^{70}-RNAP holoenzyme. (1) The holoenzyme assembles along the DNA in the closed complex. (2) Transcription initiation occurs when aromatic residues in the σ_2 domain stabilize the melting of the -10 DNA element. (3) The holoenzyme forms the open complex with DNA and RNA is transcribed along the template. (4) Abortive initiation ends and transcription can proceed beyond the initial site. (5) The elongating RNA strand pushes the σ_4 domain out of the way and causes σ^{70} to dissociate from core RNAP. (6) Elongation of the RNA transcript continues beyond the promoter region. Figure adapted from (Murakami and Darst 2003).

Though their domain architectures, sequences, and regulatory mechanisms differ, σ^{70} and σ^{54} ultimately perform the same essential function by melting the correct sequence of DNA in the promoter regions upstream of genes. While the well-studied mechanism of σ^{70} transcription initiation is likely to share similarities with that of σ^{54}, it is how σ^{54} differs in its activation requirement that will be the primary focus of this dissertation.

1.4 Structural overview of σ^{54}

Early biochemical studies divided *E. coli* σ^{54} into three regions: Region I (1-56), Region II (57-107) and Region III (108-477) (Buck et al. 2000). However, these divisions make less sense in light of newer studies that characterize a number of functionally distinct domains separated by flexible linkers. In this dissertation, we'll use a different nomenclature for σ^{54} domains that better match their function. Region I will be referred to as the activator interacting domain (AID) for its role in contacting the transcriptional activators. Region II will be referred to as the linker region, a flexible stretch of amino acids that varies in length and sequence across different species. Region III will be further divided into three domains: the core binding domain (CBD), which binds core RNA polymerase, and the -12 and -24 DNA binding domains (-12 and -24 DBD) which bind to the -12 and -24 promoter sequences upstream of the gene (Barrios, Valderrama, and Morett 1999) (Figure 1.4). Aside from the linker region, these domains are well-conserved across species including the two homologs of σ^{54} from *Escherichia coli* and *Aquifex aeolicus* used in experiments throughout this dissertation (Figure 1.5).

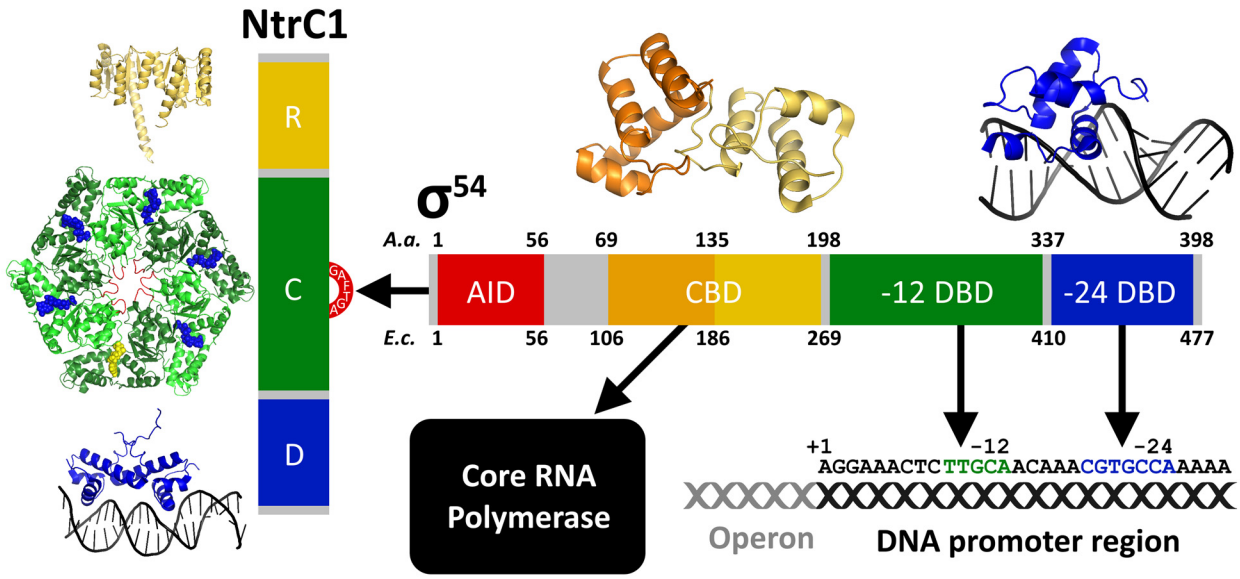


Figure 1.4 Diagram showing solved, high resolution structures of σ^{54} domains and the transcriptional activator $NtrC1^{C}$. Arrows show implicated binding events: the activator interacting domain (AID) binding to the GAFTGA loops on the central ATPase domain (C) of NtrC1, the core binding domain (CBD) binding to RNA polymerase, the DNA binding domains binding to their respective DNA elements 12 and 24 basepairs upstream of the transcription start site. The NtrC1 N-terminal receiver domain (R) regulates the assembly of the active hexamer and the NtrC1 C-terminal DNA binding domain (D) binds to enhancer sequences ≈100 bp upstream of the transcription start site.

Figure 1.5 Sequence alignment of *E. coli* σ^{54} and *A. aeolicus* σ^{54}. The secondary structure (black, coil; blue, strand; red, helix) was calculated by PSIPRED prediction. Rough domain boundaries from the five domains are marked with colored highlights (red, activator interacting domain; orange, linker region; yellow, core binding domain; green, –12 DNA binding domain; blue, –24 DNA binding domain).

A number of low resolution cryo-EM structures of full length σ^{54} and the transcriptional activator PspF have been published, which provided information about the σ^{54}-PspF complex (Rappas et al. 2005) and the full σ^{54}-RNAP-PspF complex (Bose et al. 2008). A more recent crystal structure provided a higher resolution structure of full length σ^{54} in complex with core RNA polymerase, but without DNA or the transcriptional activator (Yang et al. 2015) (Figure 1.6). These structures with full length σ^{54} help assign the locations and roles of each of the domains in the holoenzyme. Much of our existing knowledge comes from structures and biochemical assays done on isolated domains. The next sections will introduce the previous work studying structures and functions of each of the σ^{54} domains.

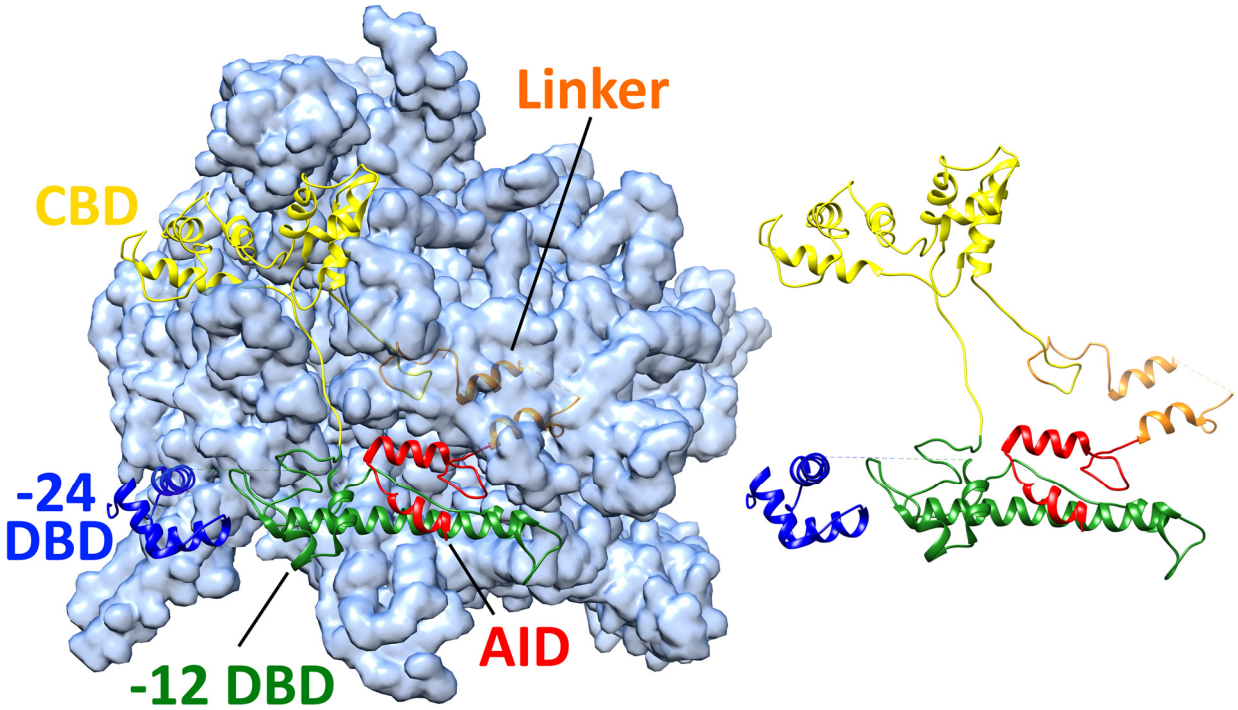


Figure 1.6 Crystal structure of σ^{54} RNAP holoenzyme. RNAP subunits α, α, β, β', and ω are shown as a surface model (transparent light blue). The subunits of σ^{54} are shown as ribbons bound to the RNAP and are from N- to C-terminus the AID (red), linker region (orange), CBD (yellow), -12 DBD (green), and -24 DBD (blue). On the right is the same structure showing only σ^{54} with the core RNAP surface model removed. Residues could not be assigned before the AID (M1 to A14), within the linker region (D71 to D80), and between the -12 and -24 DBDs (Q387 to A415). PDB ID: 5BYH (Yang et al. 2015).

1.4.1 Activator interacting domain

The σ^{54} activator interacting domain (AID) (*E. coli* residues 1-56) contacts the transcriptional activators and this interaction drives the formation of the open complex. A significant open question in σ^{54} transcription initiation is how this interaction at the N-terminus translates into a conformational change at the distant -12 and -24 DNA binding domains and the RNAP holoenzyme that makes it capable of melting DNA. One proposed mechanism suggests that the AID binds to the -12 DBD and blocks the template DNA strand from entering the active site where it would be melted by the holoenzyme (Yang et al. 2015). In this model, the transcriptional activator's interaction with the AID remodels or removes the AID from the site of DNA melting thus allowing the -12 DBD to melt the DNA. However, AID deletion mutants are defective in both activator binding and *in vivo* transcription activity (Wang, Syed, and Gralla 1997), indicating that the AID does more than block the active site. The mechanism we propose and support throughout this dissertation suggests that the AID's interaction

with the activators provides a force that drives a reconfiguration of the holoenzyme into a structure that can form the open complex formation.

The σ^{54} AID in all species has two predicted helices rich in glutamine, glutamate and leucine residues (Hsieh, Tintut, and Gralla 1994)(Hsieh and Gralla 1994). A recent crystal structure of the σ^{54}-RNAP holoenzyme fit the AID density into the structure of the holoenzyme near the -12 DBD (Yang et al. 2015). However, given the low resolution of the structure (3.76Å), the high R factor (R_{Free}=0.405), and our own work on this domain, we have reason to doubt that this density represents a biologically relevant structure of the AID. In Chapter 2, we present studies of the activator interacting domain and its interaction with NtrC1 in the ATP state. We show that the AID is an intrinsically disordered domain that binds tightly to the transcriptional activator pore loops.

1.4.2 Linker region

In between the AID and CBD lies a variable length and poorly conserved linker region (Southern and Merrick 2000). In. *A.a.* σ^{54} it is only ≈13 amino acids (57-69), while in *E.c.* σ^{54} it is closer to 50 amino acids (57-107). None of the σ^{54} linker regions have predicted secondary structure, though a recent crystal structure of *E. coli* σ^{54} found two helices at the start of the linker region (57-85) (Yang et al. 2015). From its location in the crystal structure, the authors conclude that the linker region blocks the channel that is later occupied by the single stranded DNA template during transcription, and after the AID interacts with the transcriptional activators it moves the linker region out of the channel where it no longer inhibits transcription. This proposed role for the linker region is difficult to reconcile with its lack of significant sequence homology and inconsistent length across species. Our work presented in chapter 2 suggests that this linker is entirely disordered and our work presented in chapter 3 indicates that changing its length has little effect on σ^{54} function *in vivo* but its lack of well-folded tertiary structure is required for σ^{54} function.

1.4.3 Core binding domain

The σ^{54} core binding domain (CBD) (*E. coli* residues 106-269) has some of the same properties as the σ^{70} σ_2 domain (*T. thermophilus* 74–254), which folds into eight helices and has a major surface of contact with core RNA polymerase (Vassylyev et al. 2002). Like the σ^{70} σ_2 domain, the σ^{54} CBD is negatively charged and also contacts the core

RNAP in both cryo-EM density (Bose et al. 2008) and the crystal structure (Yang et al. 2015) of the holoenzyme.

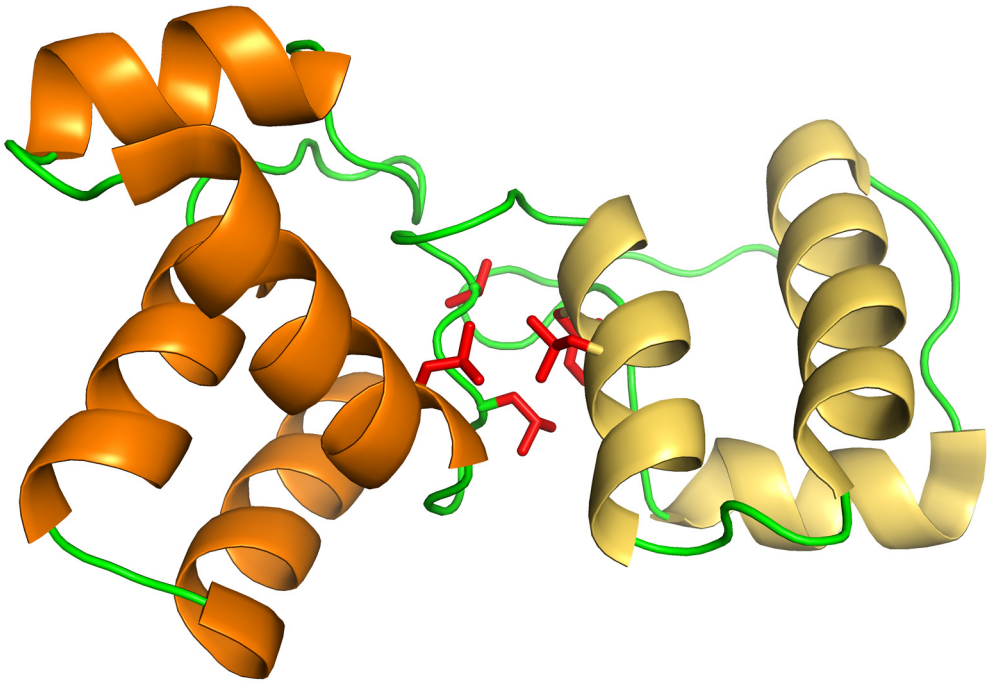


Figure 1.7 NMR Structure of the *A.a.* σ^{54} core binding domain, PDB: 2K9M (Hong, Doucleff, and Wemmer 2009). The four-helix bundle (orange) remains well-folded without the three-helix bundle (yellow). The three-helix bundle only folds against the four-helix bundle. The hydrophobic interface (L131, L137, V138, L150) between the two subdomains is shown in red and may form a conformational fracture point in the structure.

The NMR structure of the *A.a.* core binding domain (69-198) (Figure 1.7) solved in our lab folds as seven helices (Hong, Doucleff, and Wemmer 2009). It also consists of two subdomains: a four-helix bundle (60-135) and a three-helix bundle (135-198) with a hydrophobic interface between the two. In the absence of the three-helix bundle, the four-helix bundle remains well folded and with no changes to its structure. However, the three-helix bundle requires the surface of the four-helix bundle to fold against and does so even when the two subdomains are separate molecules. These two subdomains, held together by a hinge, may serve as a conformational fracture point that converts the activator binding event at the N-terminal AID into a conformational change in the C-terminal DNA binding domains that could drive the RNA polymerase holoenzyme to form the open complex. The possibility of force alone driving a conformational change in the core binding domain and leading to DNA melting is explored in Chapter 4.

1.4.4 -12 DNA binding domain

The structure of the amino acids between the core binding domain and the -24 DNA binding domain, which we will refer to as the -12 DNA binding domain (-12 DBD), is implicated in binding to and melting DNA at the -12 promoter element (Wong, Gralla, and Tintut 1994). The recent structure of the RNA polymerase holoenzyme bound to σ^{54} has shed some new light on the -12 DBD. In the structure, it consists of an extra-long helix (ELH) (*E.c.* residues 315-353) that binds to the AID, followed by a helix-turn-helix (HTH) (*E.c.* residues 365-385) that binds to the -12 DNA promoter region (Yang et al. 2015). However, given its significant amount of missing density, low resolution (3.76Å), high R-factor (R_{Free}=0.405), and our work on the AID, we have some reason to doubt that the ELH binds the AID tightly in solution.

1.4.5 -24 DNA binding domain

The structure of the *A.a.* σ^{54} -24 DNA binding domain was solved in our lab by NMR spectroscopy both free and in complex with the -24 promoter sequence (Doucleff, Malak, et al. 2005)(Doucleff et al. 2007). Another NMR structure of the *E.c.* σ^{54} -24 DNA binding domain was also solved (PDBID: 2MT3). The σ^{54} -24 DBD contains the RpoN box motif (*A.a.* 377-386) which is inserted into the major groove of the DNA and makes a number of sequence-specific contacts to bases in the -24 DNA element. Although there is no detectable sequence homology, the DNA bound structure of -24 DBD aligns well with the σ^{70} σ_4 domain that binds to the -35 DNA element in the σ^{70} promoter sequence (Ross et al. 1993) (Figure 1.8). The σ^{54} -24 DBD allows the holoenzyme to recognize the correct promoter sequence in front of genes regulated by σ^{54}. It positions the rest of the holoenzyme in the correct places to melt the DNA and initiate transcription.

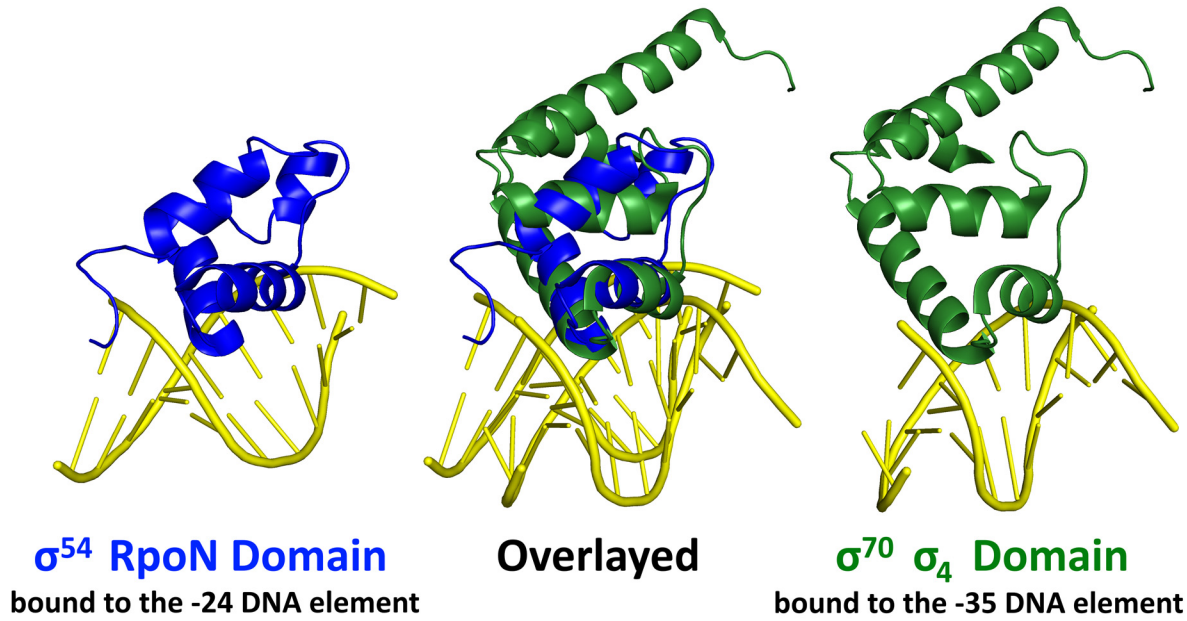


Figure 1.8 Left: The σ^{54} RpoN Domain also called the -24 DBD (blue) bound to the -24 DNA promoter element (yellow). Right: The σ^{70} σ_4 Domain (green) bound to the -35 DNA promoter element (yellow). Middle: An overlay of the two structures showing the similarity between both of these DNA binding domains. PDB ID: 208K (Doucleff et al. 2007) and 4YLO (Zuo and Steitz 2015).

1.5 Structural overview of the transcriptional activators

The σ^{54} transcriptional activators, also referred to as enhancer binding proteins (EBPs), are a class of proteins with a conserved AAA+ ATPase domain that activate σ^{54}-mediated transcription initiation of specific genes. Typically, two pairs of dimeric transcriptional activators bind to two enhancer sequences roughly 100 basepairs upstream of the transcription start site using a sequence specific DNA binding domain (D) (Wedel et al. 1990b). They are activated through their N-terminal regulatory domain (R), often by phosphorylation, which promotes oligomerization of the central ATPase domain (C) (De Carlo et al. 2006), generally into hexamers (Batchelor et al. 2009). The active hexameric transcriptional activators contact the N-terminal activator interacting domain (AID) of σ^{54} and remodel the σ^{54}-RNAP holoenzyme through rounds of ATP hydrolysis (Wedel and Kustu 1995). The structural changes induced in the holoenzyme drive melting of the DNA at the -12 promoter site and open complex formation, thereby initiating transcription of the gene.

1.5.1 Regulatory domain

The regulatory domains respond to environmental signals to control the timing of the transcriptional activation of σ^{54}. There are a large variety of different types of regulatory domains including the NifA GAF domain, the DhaR PAS domain, and the NtrC receiver domain discussed in more detail throughout this dissertation. Some transcriptional activators use multiple regulatory domains and others, like PspF, lack any regulatory domain and are controlled by a second protein (Studholme and Dixon 2003). There are structures of a number of activated and unactivated regulatory domains including DctD (Park et al. 2002), NtrC (Kern et al. 1999)(Hastings et al. 2003) (Figure 1.9), NtrC1 (Lee et al. 2003)(Doucleff, Chen, et al. 2005), NtrC4 (Batchelor et al. 2008), and NifA-like homologs Nlh1 and Nlh2 (Batchelor et al. 2013).

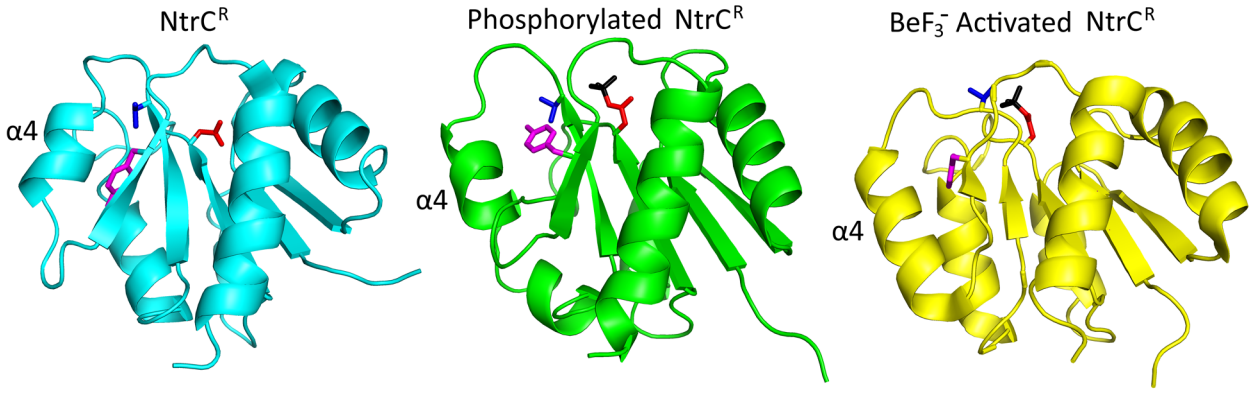


Figure 1.9 Ribbon representation of apo-NtrC^R (left, 1DC7), phosphorylated NtrC^R (middle, 1DC8), and BeF_3^--NtrC^R (right, 1J56). Side chains of Asp54 (red), Thr82 (blue), and Tyr101 (magenta) are shown. The phosphate and BeF_3^- attached to Asp54 are shown in black. The fourth α-helix, which undergoes the most significant rearrangement when NtrC^R is activated, has been labeled.

The σ^{54} transcriptional activators used throughout most of the work presented in this dissertation come from the NtrC family. For NtrC, the receiver domain is phosphorylated on a conserved aspartate by a histidine kinase, NtrB, at the end of two component signal transduction (Keener and Kustu 1988). Phosphorylation can be part of a negative regulatory mechanism if it alters the receiver domain dimer interface allowing NtrC-like proteins to form a hexamer as in *Aquifex aeolicus* NtrC1 (Doucleff, Chen, et al. 2005) or a positive regulatory mechanism if it induces a conformational change in the receiver domain that allows it to bind the ATPase domain of an adjacent protein promoting the oligomerization of the hexamer as in *E. coli* NtrC (De Carlo et al. 2006). Some transcriptional activators in the NtrC family fall in between positive and negative regulation, as is the case with *A.a.* NtrC4. This dissertation goes into the

details of NtrC-like receiver domain activation in Chapter 5, where we also characterize an NtrC receiver domain from the piezophilic bacteria *Shewanella violacea*.

1.5.2 AAA+ ATPase domain

The AAA+ ATPase domain is highly conserved among all transcriptional activators (Zhang et al. 2002). The ATPase domains often bind the enhancer DNA as inactive dimers, as in the case of negatively regulated NtrC1 and to a lesser extent NtrC4 (Batchelor et al. 2008), and only oligomerize into active hexamers once their regulatory domain receives a signal.

The *A.a.* NtrC1 AAA+ ATPase domain was first crystallized as heptamers when all subunits were trapped in the same nucleotide state; either with ADP, with the non-hydrolyzable ATP analog ADP-BeF_3^- (Lee et al. 2003), or with ATP using the mutant ATPase domain $NtrC1^C$(E239A) which lacks the ability to hydrolyze ATP (Chen et al. 2010). However, mass spectrometry and SAXS experiments revealed that the hexamer is the predominant structure in solution and suggests that heptamer is an artifact of trapping the structure with an unnatural, uniform arrangement of ATP or ADP during crystallography (Batchelor et al. 2009). Crystal structures of other activators include the hexameric ATPase domain of PspF in complex with ATP analogs (Rappas et al. 2006), the hexameric ATPase and DNA binding domains of ZraR (Sallai and Tucker 2005), and an EM structure of NtrC (De Carlo et al. 2006). A recent crystal structure of the NtrC1 ATPase domain in a mixed nucleotide state may more closely resemble the structure of a typical ATPase domain inside the cell (Sysoeva et al. 2013). The mixed state folds into an asymmetric ring with a lock-washer structure where the 1^{st} and 6^{th} subunits do not form many contacts. This is similar to the structure of some other unrelated AAA+ ATPases involved in different cellular process (Neuwald et al. 1999), in particular the Rho helicase (Skordalakes and Berger 2006), and to a lesser extent the ClpX subunit of the ClpXP protease (Glynn et al. 2009) and the DnaB helicase (Thomsen and Berger 2009) which fold as hexamers around a central pore, though they do not form asymmetric lock-washers.

An important feature of the oligomeric AAA+ ATPase domains is a short loop found at the entrance to pore, which consists of a highly conserved "GAFTGA" motif. Mutations to these GAFTGA loops prevent the transcriptional activator from binding to σ^{54} (Chaney et al. 2001)(Dago et al. 2007). The ATP binding site is located between adjacent subunits of the ring. The ATP binding cleft (P-loop and Walker B motif) of one subunit contacts the nucleotide (ATP or ADP) which forms a bridge to the arginine

finger of the adjacent subunit (Chen et al. 2010)(Sysoeva et al. 2013). Therefore, the occupancy of the nucleotide binding site (apo, ADP, or ATP) affects the structure of the subunit, its neighbor, and the organization of the entire oligomer. When a subunit is bound to ATP, its GAFTGA loop protrudes further above the pore, but when a subunit is bound to ADP or no nucleotide its GAFTGA loops retract toward the pore (Figure 1.10). Along with the biochemical data showing that $NtrC1^{C}$ can only bind to σ^{54} when it is trapped in its ATP state (Chen et al. 2007), this suggests that the protrusion of the GAFTGA loops exposes a binding surface recognized by the σ^{54} AID.

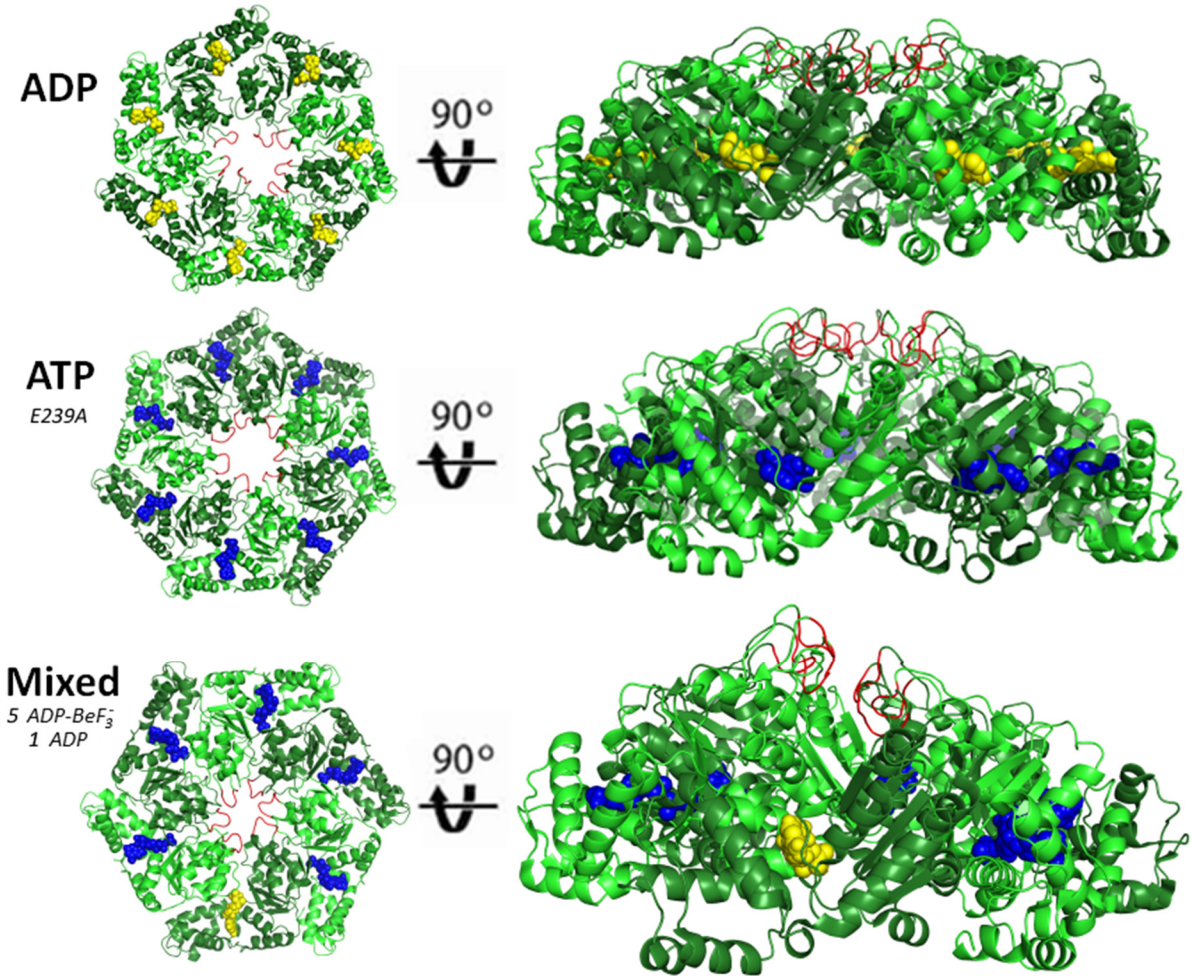


Figure 1.10 Crystal structures showing the $NtrC1^{C}$ heptamer (green) uniformly occupied with ADP (yellow), the ATP hydrolysis deficient mutant $NtrC1^{C}$(E239A) heptamer uniformly occupied with ATP (blue), and the $NtrC1^{C}$ hexamer bound to a mixture of nucleotides ADP (yellow) and ADP-BeF_3^- (blue). The location of the conserved GAFTGA loops are highlighted (red). A side view of each structure shows the raising of the GAFTGA loops in the ATP-like nucleotide state. PDB from top to bottom: 1NY6, 3M0E, 4LY6.

Contact between σ^{54}'s activator interacting domain and the transcriptional activator GAFTGA loops of subunits in the ATP state is not sufficient to activate σ^{54}. At least one round of ATP hydrolysis by the transcriptional activators is required to drive the holoenzyme into the open complex (Chaney et al. 2001). There is no clearly understood mechanism to explain how the σ^{54}-transcriptional activator interaction leads to conformational changes in the σ^{54}-RNAP holoenzyme or how these conformational changes enable the holoenzyme to melt DNA. A central goal of this dissertation to better understand the activation process by studying the initial σ^{54}-transcriptional activator binding event, the ATP hydrolysis driven activation mechanism, and how activation drives conformational changes to the remainder of the holoenzyme. Our work on these three topics will be discussed further in Chapters 2, 3, and 4 respectively.

1.5.3 DNA binding domain

The DNA binding domain localizes the transcriptional activators to enhancer sequences ≈100 basepairs upstream of the transcription start site (Wedel et al. 1990a) (Figure 1.11) Typically, as is the case with NtrC, there are two sets of two binding sites that allow two transcriptional activator dimers to assemble right next to one another along the DNA. These two dimers oligomerize with a third dimer that comes in from solution to complete the hexamer (Batchelor et al. 2009). Some σ^{54} transcriptional activators like NorF require three binding sites (Tucker et al. 2009), while others like *Helicobacter pylori* FlgR have no recognizable DNA binding domain and apparently assemble and activate σ^{54} from solution (Brahmachary et al. 2004).

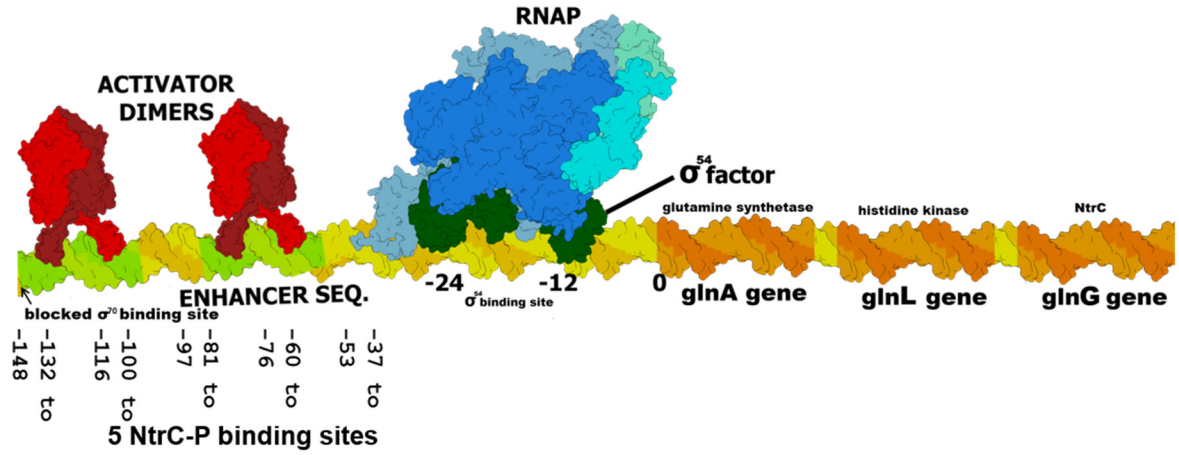


Figure 1.11 Cartoon showing a typical σ^{54}-regulated operon, glnALG. Two NtrC dimers (red) bind to four out of five detected NtrC enhancer sites (lime green) in front of the σ^{54} (green) and RNA polymerase holoenzyme (blue) bound at the conserved -24 and -12 site along the DNA (yellow). After melting the DNA it begins transcription of the glnA gene (orange) as well as genes coding for the histidine kinase that activates NtrC and NtrC itself.

Recently, the structure of the NtrC1 DNA binding domain was solved by X-ray crystallography in our lab (Vidangos et al. 2013) (Figure 1.12). The DNA binding domains contact the enhancer sequences as dimers. A helix turn helix motif is inserted into the major groove forming sequence specific contacts with the enhancer sequences and inducing a slight bend to the DNA. A flexible region between the central ATPase domain and the DNA binding domain allows a structural rearrangement of the central ATPase domain once the protein has been activated at the receiver domain.

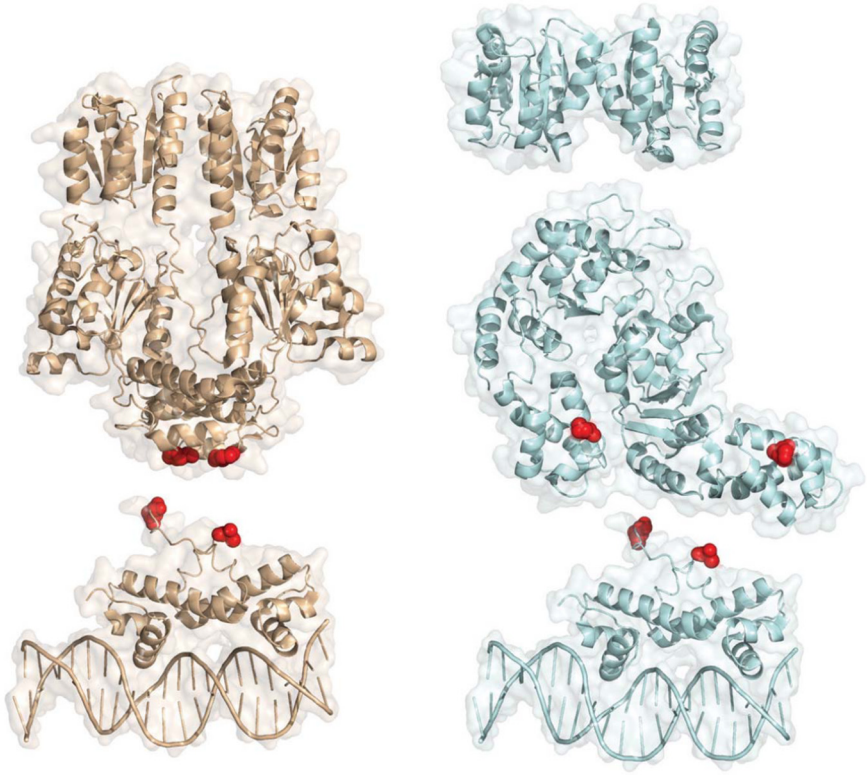


Figure 1.12 Model of the activation of NtrC1 bound to DNA enhancer sequences. The structure of the NtrC4 "off" state (left, orange) where the receiver and ATPase domains form inactive dimers. The structure of the activated NtrC4 "on" state (right, blue) where phosphorylation of the receiver causes the receiver and ATPase domains to change their dimer interfaces leading to oligomerization of the active hexamer. The flexible linker between the DNA binding domain and central domain (ends highlighted in red) accommodates the rearrangement of the dimers upon activation. Figure adapted from (Vidangos et al. 2013).

Two important roles of the DNA binding domain are to promote the hexamerization of the ATPase domain and the interaction between transcriptional activators and the σ^{54}-RNAP holoenzyme. While $NtrC^{3Ala}$ mutants with a defective DNA binding domain are 5000-fold less active *in vitro*, their activity can be recovered by overexpressing NtrC, promoting its oligomerization in solution (Porter et al. 1993)(North and Kustu 1997). Furthermore, NtrC is still able to activate σ^{54} transcription when the enhancer binding

sites are shifted 1000 basepairs upstream from the rest of the promoter (Schulz, Langowski, and Rippe 2000) or when the enhancer sites and σ^{54} promoter elements are on different DNA strands of two concatenated plasmids (Wedel et al. 1990a). This indicates that enhancer binding is needed to increase the local concentration of the activators to promote their cooperative oligomerization into active hexamers and the proximity of their binding sites to the σ^{54} promoter region increases the frequency of their collisions with σ^{54}.

1.6 Mechanism of Transcriptional Activation of σ^{54}

A major open question in σ^{54} transcriptional activation is how the interaction between the σ^{54} activator interacting domain and the transcriptional activators drives remodeling of the RNAP holoenzyme to form the open complex with DNA. Two cryo-EM density maps, one showing the structure of the transcriptional activator PspF bound to σ^{54} (Rappas et al. 2005) and the other showing the PspF bound to the full σ^{54}-RNAP holoenzyme (Bose et al. 2008), identified four lobes of σ^{54} density which roughly correspond to the AID, CBD, -12 DBD and -24 DBD. In the structure without PspF, the σ^{54} AID appears to be blocking the DNA from accessing the active site where the -12 DBD would melt it. When PspF binds σ^{54}, there is a movement of the AID density that suggests it no longer blocks the active site. A more recent crystal structure of the σ^{54}-RNAP holoenzyme (Yang et al. 2015) also mapped the AID into a position bound to the -12 DBD close to its active site, though no high resolution structures of σ^{54} bound to a transcriptional activator have been published.

The structural evidence from the holoenzyme suggests that contact between the transcriptional activators and the σ^{54} AID pulls it away from the active site of RNAP where DNA melting occurs and that this is sufficient to promote the formation of the open complex (Yang et al. 2015). This mechanism is supported by results showing that σ^{54} lacking the AID can initiate transcription of pre-melted 'bubble' DNA though it cannot open double stranded DNA (Wang, Syed, and Gralla 1997) and that mutations to specific leucines in the AID (*E. coli* σ^{54} L26S,L33S) allow full length σ^{54} to initiate transcription without the activator *in vitro* (Wang et al. 1995). However, it is hard to reconcile this mechanism with some of the biochemical evidence. In particular, σ^{54} mutants with a truncated AID cannot initiate transcription, which they would be expected to if their only purpose was blocking the site of DNA melting before

activation. Additionally, contact between σ^{54} and the GAFTGA loops of subunits trapped in the ATP state with a non-hydrolyzable ATP analog is not sufficient to form an open complex (Chaney et al. 2001). The ATPase domain must undergo at least one round of ATP hydrolysis to activate σ^{54} and melt the DNA. This opens up the possibility that the mechanism of activation is more dynamic than structural snapshots of the holoenzyme can capture.

In Chapter 3, we propose an alternative mechanism to the simple removal of the inhibitory AID that has been hypothesized based on the existing σ^{54}-RNAP holoenzyme structures. It is notable that the AAA+ ATPase transcriptional activators share a number of structural characteristics with other AAA+ ATPases (Tucker and Sallai 2007). In particular, the structure of Rho helicase is also a hexameric lock-washer, also contains a central pore lined with loops that move in response to ATP binding and hydrolysis, and those loops are also implicated in binding to Rho's target molecule, in this case RNA (Skordalakes and Berger 2006). A functional model of Rho ATPase hypothesizes that sequential ATP hydrolysis drives motion of the loops, which push RNA through the central pore of the ring. Drawing on the similarities between the structure of Rho and other AAA+ ATPases and σ^{54} transcriptional activators like NtrC1 (Sysoeva et al. 2013), we propose a threading mechanism of σ^{54} transcriptional activation. In this model mechanism, force is applied to σ^{54} by the motion of the GAFTGA loops during ATP hydrolysis, which threads the flexible N-terminal σ^{54} AID through their central pore. This force drives a conformational change in σ^{54} and the RNA polymerase holoenzyme, perhaps through a conformational fracture point in the σ^{54} core binding domain. The reconfigured RNAP holoenzyme is then capable of melting the DNA at the -12 promoter element and forming the open complex.

1.7 Two component signal transduction

σ^{54} transcriptional activation by NtrC is part of a two-component signal transduction pathway. There are many different two-component systems (TCSs) in nearly all bacterial species, including over 62 in *E. coli* alone (Mizuno 1997). TCSs trigger a cellular response when they sense environmental stimuli such as changes in temperature, pressure, pH, cation concentrations, nutrients, oxygen content and others (Beier and Gross 2006). TCSs require at least two proteins, a sensor histidine kinase (HK) and a response regulator (RR). When the HK senses a specific environmental stimulus it transmits the signal to the RR, which initiates the cell's response (West and Stock 2001).

HKs are multidomain proteins with sensory and kinase domains. A conserved histidine in the kinase domain is autophosphorylated and then transfers its phosphate to a conserved aspartate in the RR to activate it (Keener and Kustu 1988). Many HKs also have phosphatase activity and can dephosphorylate a phosphoaspartate on the response regulator to deactivate it (West and Stock 2001). The vast majority (96%) of HK sensory domains are activated by small molecule binding, with the remaining few activated by protein–protein interactions. The RR generally consists of at least two domains, a receiver domain that receives the phosphate from the HK and is activated, and an output domain that triggers the cell's response to the input signal. The output domain of the majority (87%) of response regulators is a DNA-binding helix-turn-helix that either inhibits or stimulates transcription of target genes, with the remaining outputs controlling levels of cyclic nucleotides or protein phosphorylation (Ulrich, Koonin, and Zhulin 2005).

Two-component signaling pathways are not as common as one-component signaling pathways in bacteria and are likely a more recent evolutionary adaptation (Ulrich, Koonin, and Zhulin 2005). A majority (83%) of HKs include transmembrane helices and are bound to the membrane. This may explain the evolutionary advantage of adding a second component to signaling pathways as it allows membrane bound HKs to sense environmental stimuli outside the cell and transmit that signal into the cell by activating a cytosolic response regulator. Since the second component is free to travel throughout the cell, it can move away from the membrane and bind DNA to regulate transcription. NtrB is one of the few (17%) cytosolic HKs that respond to an internal signal rather than an external one by sensing intracellular levels of nitrogen. Like other HKs, NtrB transfers its phosphate to the response regulator NtrC (Keener and Kustu 1988)(Weiss and Magasanik 1988), which has a helix-turn-helix domain that binds to enhancer sequences ≈100 basepairs upstream of transcription start sites (Pelton, Kustu, and Wemmer 1999) to express proteins that help the cell survive on limited nitrogen (Ninfa, Reitzer, and Magasanik 1987)(Reitzer 2003). Chapter 5 goes into more detail on the NtrB/NtrC two-component signal transduction pathway and describes our studies of an NtrB/NtrC pathway in a piezophilic bacteria, *Shewanella violacea*, that senses and responds to high pressure.

1.8 Dissertation Outline

This dissertation spans a variety of topics relating to σ^{54} transcription initiation, primarily focusing on its activation by the transcriptional activators in the NtrC family. Chapter 2 describes our work characterizing the structure of the *Aquifex aeolicus* σ^{54} activator interacting domain (also known as Region I) and its ATP-dependent binding with the transcriptional activator NtrC1. Chapter 3 describes our work studying the mechanism of *Escherichia coli* σ^{54} activation by the transcriptional activator NtrC through *in vivo* transcription assays using insertions and deletions to the amino acid sequence in or around the σ^{54} activator interacting domain. The results included in chapters 2 and 3 are in preparation for submission to peer-reviewed journals. Chapter 4 describes a side project that uses molecular tweezers to study a hypothesized force-dependent σ^{54} activation mechanism by looking for conformational changes to the *Aquifex aeolicus* σ^{54} core binding domain under mechanical force. Chapter 5 describes a side project using high pressure NMR spectroscopy to study the regulatory receiver domain of a piezophilic transcriptional activator *Shewanella violacea* NtrC and how the cell's two-component signal transduction pathway controls the cell's response to pressure as an environmental signal.

Chapter 2: Role of the σ^{54} Activator Interacting Domain in Bacterial Transcription Initiation

2.1 Summary

Bacterial sigma factors are subunits of RNA polymerase that direct the holoenzyme to specific sets of promoters in the genome, and are a central element of regulating transcription. Most polymerase holoenzymes open the promoter and initiate transcription rapidly after binding. However, polymerase containing members σ^{54} family must be acted on by a transcriptional activator before DNA opening and initiation occur. A key domain in these transcriptional activators forms a hexameric AAA+ ATPase that acts through conformational changes brought on by ATP hydrolysis. Contacts between the transcriptional activator and σ^{54} are primarily through the N-terminal σ^{54} activator interacting domain (AID). To better understand this mechanism of bacterial transcription initiation, we characterized the σ^{54} AID by NMR spectroscopy and other biophysical methods, and show that it is an intrinsically disordered domain in σ^{54} alone. We identified a minimal construct of the *A. aeolicus* σ^{54} AID that consists of two predicted helices and retains native-like binding affinity for the transcriptional activator NtrC1. Using the NtrC1 ATPase domain, bound with the non-hydrolyzable ATP analog ADP-beryllium fluoride, we studied the NtrC1-σ^{54} AID complex using NMR spectroscopy. We show that the σ^{54} AID becomes structured after associating with the core loops of the transcriptional activators in their ATP state and that the primary site of the interaction is the first predicted helix. Understanding this complex, formed as the first step toward initiation, will help unravel the mechanism of σ^{54} bacterial transcription initiation.

2.2 Introduction

The five subunits of the bacterial "core" RNA polymerase, α, α, β, β' and ω, are sufficient for transcribing mRNA once the promoter has been opened. However, in order to recognize promoter sequences, 'melt' the promoter DNA, and initiate transcription, the core RNA polymerase requires an additional, modular subunit, the sigma factor

(Burgess et al. 1969). The sigma factors bind to the core RNA polymerase, forming the RNA polymerase holoenzyme. The sigma factors bind sequence specifically to DNA in the promoter region, with different sigma factors targeting different subsets of genes to accomplish differential transcriptional regulation (Losick and Pero 1981). Much of the regulation occurs by controlling the formation of the promoter-holoenzyme complex, either through anti-sigma proteins that compete with polymerase for a particular sigma factor (Campbell, Westblade, and Darst 2008), or through repressors that block the promoter (Colland et al. 1998). Once the RNA polymerase holoenzyme-promoter complex forms, the sigma factors help open DNA and initiate transcription. After initiation the sigma factor can disassociate from the complex and the core RNA polymerase can continue to transcribe mRNA using the single stranded DNA template (Liu and Martin 2002)(Tahirov et al. 2002) (Figure 2.1).

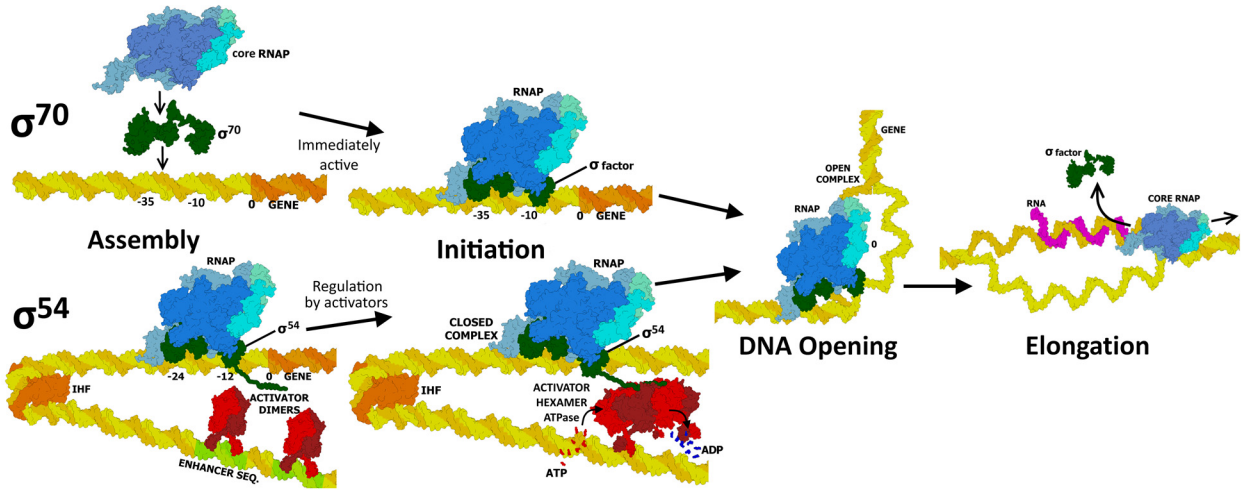


Figure 2.1 Diagram of transcription initiation mediated by the two classes of bacterial sigma factors. (1) Assembly: sigma factor and RNA polymerase assemble upstream of the start site. (2) Initiation: σ^{70} is immediately able to initiate transcription, while σ^{54} requires an activation event from one of the transcriptional activators. (3) DNA opening: the RNA polymerase holoenzyme melts DNA. (4) Elongation: the sigma factor falls off and core RNA polymerase continue to transcribe RNA off the single stranded DNA.

Sigma factors have been divided into two broad families that share no sequence homology: the more common σ^{70} family and the rarer σ^{54} family (Buck et al. 2000)(Chaney et al. 2001). All sigma factors serve the same purpose in directing RNA polymerase to specific promoters, but they differ in their mechanism of action and regulation. All sigma factors bind to core RNA polymerase to form a holoenzyme and all bind promoter regions slightly upstream from the transcription start site. σ^{70} RNA polymerase holoenzyme is capable of opening promoter DNA and initiating transcription immediately after binding the promoter (Campbell et al. 2002). However, σ^{54}

polymerase requires an additional activation step, a conformational change that is driven by a transcriptional activator, before it can open the promoter (Buck et al. 2000). The σ^{54}-RNAP holoenzyme recognizes conserved sequences –24 and –12 basepairs upstream of the transcription start site (Barrios, Valderrama, and Morett 1999) where it binds and awaits activation by a transcriptional activator that assembles further upstream (Sasse-Dwight and Gralla 1988). The transcriptional activators themselves must be triggered, often in response to an environmental stimulus (Keener and Kustu 1988), after which they act on the σ^{54}-RNAP holoenzyme, which then transcribes the DNA for the encoded protein initiating gene expression (Kustu et al. 1989). The additional activation requirement affords genes under control of σ^{54} an extra layer of control that both reduces background levels of transcription and gives a rapid cellular response when conditions are right. Consistent with this behavior, genes regulated by σ^{54} include those necessary for response to starvation and heat shock among others (Kazmierczak, Wiedmann, and Boor 2005). The detailed mechanism by which these transcriptional activators reconfigure σ^{54} and the RNAP holoenzyme into a form capable of opening DNA is not known.

The σ^{54} transcriptional activators typically have three functional domains: (1) an N-terminal regulatory domain that receives a signal and promotes assembly of the active, hexameric form of the activator; (2) a central AAA+ ATPase domain that binds σ^{54} and hydrolyzes ATP; and (3) a C-terminal DNA binding domain that binds to enhancer sequences well upstream of the site of DNA melting (Rappas, Bose, and Zhang 2007). Regulatory domains are quite diverse (Nixon, Ronson, and Ausubel 1986)(Ronson, Nixon, and Ausubel 1988)(Studholme and Dixon 2003), responding to different kinds of signals including phosphorylation of a receiver domain (Lee et al. 2003), or ligand binding by a GAF domain (Batchelor et al. 2008). Regulation can be positive, for example phosphorylation of the receiver domain promoting formation of the active hexamer ATPase, or negative where the domain inhibits formation of the hexamer until the signal is received. The central domain of negatively regulated activators may oligomerize into an active conformation when expressed without its regulatory and DNA binding domains, as is the case with the NtrC1 central domain construct ($NtrC1^{C}$) (Batchelor et al. 2009) used in the experiments reported here (Figure 2.2).

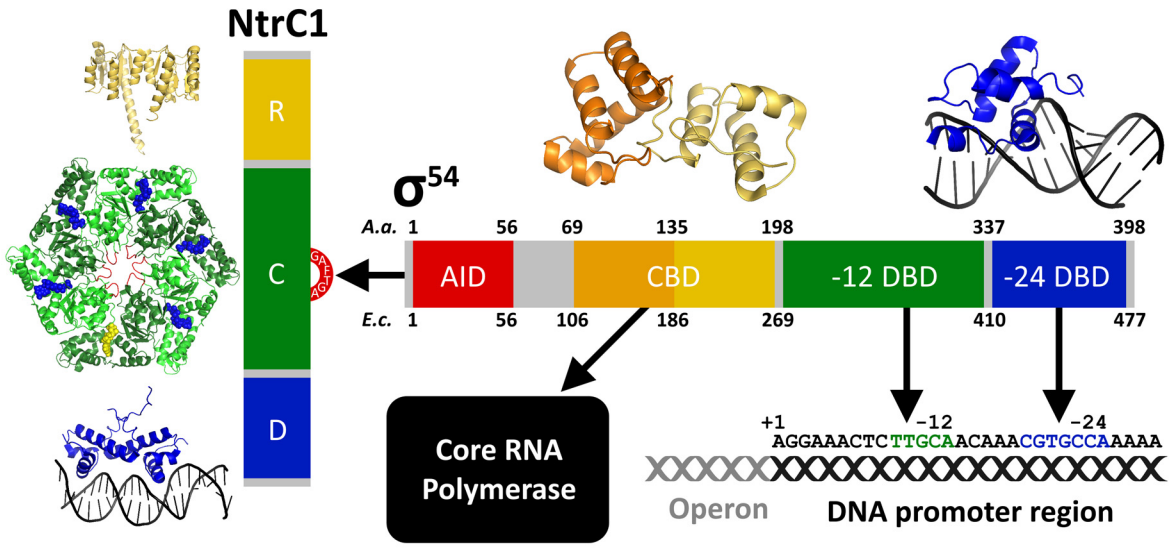


Figure 2.2 Diagram showing solved, high resolution structures of σ^{54} domains and the transcriptional activator $NtrC1^{C}$. Arrows show implicated binding events: the activator interacting domain (AID) binding to the GAFTGA loops on the central ATPase domain (C) of NtrC1, the core binding domain (CBD) binding to RNA polymerase, the DNA binding domains binding to their respective DNA elements 12 and 24 basepairs upstream of the transcription start site. The NtrC1 N-terminal receiver domain (R) regulates the assembly of the active hexamer and the NtrC1 C-terminal DNA binding domain (D) binds to enhancer sequences ≈100 bp upstream of the transcription start site.

Activated transcriptional activators assemble into hexameric rings with six ATP binding sites, each at the cleft between subunits (Lee et al. 2003). A highly conserved loop at the top of the central pore, with the sequence GAFTGA, has been shown to be involved in the interaction between the activator and σ^{54} (Chen et al. 2007)(Rappas et al. 2006). Crystal structures show that the GAFTGA loop extends upward on subunits bound to ATP (or a non-hydrolyzable ATP analog), but retract inward for subunits bound to ADP (Lee et al. 2003)(Chen et al. 2010)(Sysoeva et al. 2013).

σ^{54} has several functional domains, two of which previously had structures determined (Doucleff et al. 2007)(Hong, Doucleff, and Wemmer 2009) (Figure 2.2). The focus of the present work is the N-terminal ≈ 50 amino acids of σ^{54}, which are responsible for interacting with the assembled ATPase of the activator that we term the activator interacting domain (AID). This is followed by a variable length, low conservation linker, and then the core binding domain, which consists of two subdomains, a four-helix and three-helix bundle, that dock together (Hong, Doucleff, and Wemmer 2009). The core binding domain is a primary region of interaction with core RNA polymerase subunits, stabilizing the holoenzyme. The core binding domain was proposed as a potential conformational fracture point that could reconfigure the holoenzyme form, melting DNA in the process of activation. The next domain, which we term the –12 DNA binding domain, interacts with DNA in the –12 region of the promoter where DNA opening

occurs (Wong, Gralla, and Tintut 1994). While a consensus –12 DNA sequence has been identified, there is variability in this region of the promoter. DNA opening is initiated in the –12 region, and the part of σ^{54} in contact with it is likely responsible for helping to stabilize the opened 'bubble' that forms during transcription initiation (Oguiza et al. 1999)(Buck et al. 2006)(Bose et al. 2008). The final, C-terminal segment of σ^{54} is a helix-turn-helix sequence-specific DNA binding domain. This region binds to the strongly conserved –24 element of DNA called the RpoN box in σ^{54} driven promoters, making numerous sequence specific contacts (Doucleff et al. 2007). This interaction also fixes σ^{54} in the correct position along the DNA such that the –12 binding domain can interact with the correct region of the promoter (Figure 2.3).

Figure 2.3 Sequence alignment of *E. coli* σ^{54} and *A. aeolicus* σ^{54}. The predicted secondary structure (black, coil; blue, strand; red, helix) calculated by PSIPRED prediction is shown. Rough domain boundaries from the five domains are marked with colored highlights (red, activator interacting domain; orange, linker region; yellow, core binding domain; green, –12 DNA binding domain; blue, –24 DNA binding domain).

The mechanism by which the N-terminal activator interacting domain (AID) functions in the σ^{54} activation remains unclear. It has been shown that deletion of the first 50 residues of the N-terminal AID (Wang, Syed, and Gralla 1997), or any mutation in the conserved GAFTGA loop of the transcriptional activators (Chaney et al. 2001), prevent binding of σ^{54} to the activator *in vitro* and activation is completely lost *in vivo*. However, most single amino acid substitutions and even deletions of small regions within the AID have little effect on activator binding and activity (Pitt, Gallegos, and Buck 2000)(Gallegos and Buck 2000). σ^{54} only binds the transcriptional activators with significant affinity when the activator subunits are primarily in the ATP state (Chen et al. 2007), which should be the default state in the cell. Activator binding is not detected when only ADP is present. Thus a primary role of the AID seems to be in forming the initial complex with the assembled activator ATPase domains.

Contact between σ^{54} and the activator alone is not enough to initiate transcription, ATP hydrolysis by the activator is required before the σ^{54}-RNAP holoenzyme bound to DNA can melt the DNA and transition from a closed to an open complex (Wedel and Kustu 1995)(Chaney et al. 2001). The extent of ATP hydrolysis required for opening has not been determined. Structural information about the initial encounter complex may shed light on how ATP hydrolysis by the activator ATPase domains bound to the AID is coupled to open complex formation through structural rearrangements of other domains of the σ^{54}-RNAP holoenzyme.

Attempts to study the structure and interactions of σ^{54} have been hindered by the poor solubility of σ^{54} constructs that include the AID. Crystallization of σ^{54} alone has only been possible with the AID removed (Darst, personal communication). Recently full length σ^{54} was crystallized in the presence of RNA polymerase (Yang et al. 2015), but the low resolution of the structure makes interpretation of the AID structure difficult, and does not address how it interacts with the transcriptional activators. In the holoenzyme crystal structure, the AID appears to fold against RNAP as two helices near the –12 site of the DNA where the RNAP holoenzyme initiates DNA melting. However, here we present evidence that the σ^{54} AID is an unstructured, intrinsically disordered domain, a segment of which becomes immobilized in the initial complex with the transcriptional activator, probably contacting the activator's GAFTGA loops. We wanted to explore whether the AID serves as a flexible tether that extends out from the rest of σ^{54} so that the AAA+ ATPase transcriptional activators can bind, localizing the activator near polymerase so that subsequent ATP hydrolysis can drive conformational changes in the σ^{54}-RNAP holoenzyme that ultimately opens DNA and leads to initiation of transcription.

2.3 Results

2.3.1 Disorder in the σ^{54} N-terminal Activator Interacting Domain

Our previous structural work has been done using domains from the thermophile *Aquifex aeolicus* σ^{54} (Hong, Doucleff, and Wemmer 2009)(Doucleff, Malak, et al. 2005). Several activators from this organism have also been characterized (Lee et al. 2003)(Batchelor et al. 2008)(Batchelor et al. 2013), which encouraged us to continue studies of it. However, a majority of genetic and biochemical work has used *Escherichia coli* (or closely related organisms), and so we have also made and studied similar constructs of the *E.c.* σ^{54}. We chose to express protein constructs that correspond to just the N-terminal activator interacting domain of σ^{54}, a central fragment of it, or include the full AID and continue through neighboring domains. The longer constructs terminate with the four-helix bundle of the core binding domain, or the full core binding domain, or correspond to full length σ^{54}. Amino acid sequence alignments were used to design constructs covering the equivalent domains from the *A.a.* and *E.c.* versions of σ^{54}. Each protein was expressed in ^{15}N labeled growth medium, purified, and ^{1}H-^{15}N HSQC spectra were collected using an 800 MHz NMR spectrometer.

A.a. σ^{54}(1-135) includes the AID and the four-helix bundle of the core binding domain. HSQC spectra from this construct were compared with those from the previously studied *A.a.* σ^{54}(60-135), for which the structure was determined (Hong, Doucleff, and Wemmer 2009). Nearly all of the amide peaks in the previously assigned σ^{54}(60-135) spectrum perfectly match the peaks in the σ^{54}(1-135) spectrum (Figure 2.4). Almost all of the additional peaks in the σ^{54}(1-135) spectrum have low dispersion of chemical shifts in the ^{1}H dimension. These peaks correspond to the first 60 residues of σ^{54} and the low shift dispersion is characteristic of an unfolded protein segment. The same comparison was done with the *E.c.* core binding domain construct σ^{54}(106-269) and the longer version containing the AID, linker, and CBD, *E.c.* σ^{54}(1-269). The results are the same, the σ^{54}(106-269) peaks are well-dispersed and nearly all of them overlay with a peak in the σ^{54}(1-269) spectrum. The extra peaks in the σ^{54}(1-269) construct, which must correspond to the first 105 residues of *E.c.* σ^{54}, are poorly dispersed and in the region of the spectrum that corresponds to unfolded residues. For both the *A.a.* and *E.c.* versions there are only very minor changes in the chemical shifts of CBD peaks in the presence of the AID and linker domain, which indicates that there is no significant contact between the CBD and the AID.

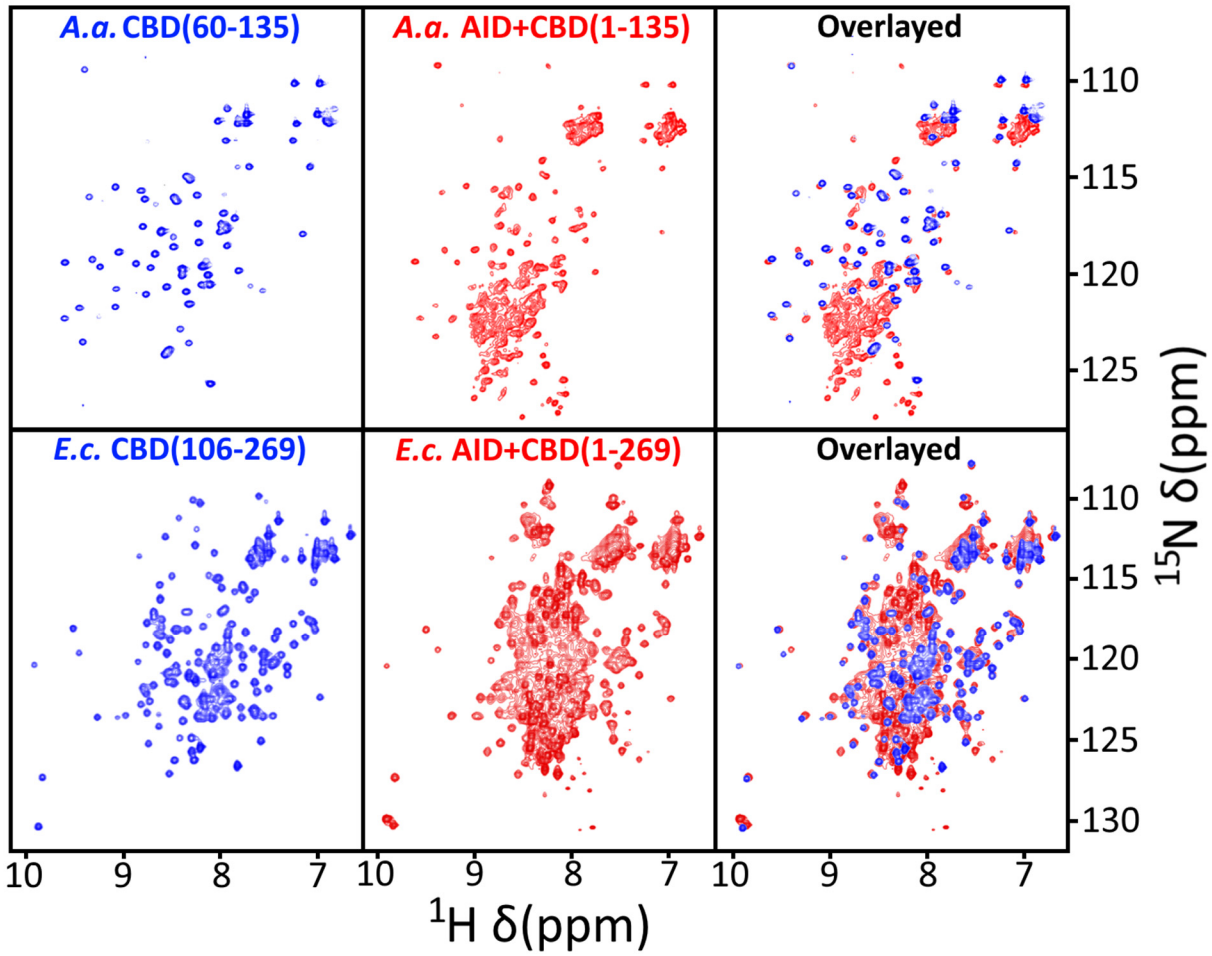


Figure 2.4 Comparison of ^{1}H-^{15}N HSQCs from the core binding domain (left, blue) and the activator interacting domain plus the core binding domain (red, middle) of *A. aeolicus* (top) and *E. coli* (bottom) σ^{54}. In both species, overlaying the HSQC of the core binding domain alone with a construct containing both the activator interacting domain and core binding domain (right) reveals unaccounted for peaks corresponding to the unstructured activator interacting domain and linker region.

Other constructs of *E.c.* σ^{54} were also studied, including the full length protein (residues 1-477), the AID-linker-4 helix bundle σ^{54}(1-186), the AID alone σ^{54}(1-62), and a segment of the AID alone σ^{54}(11-48). In all of these the pattern of a poorly dispersed peaks associated with the AID, with ^{1}H shifts between 8 and 9 ppm, occurs (Figure 2.5). The consistently low dispersion of AID peaks and the lack of change in the chemical shifts of well-folded residues in other σ^{54} domains in the presence of the AID show that the AID in free σ^{54} is unstructured, and does not interact significantly with other domains.

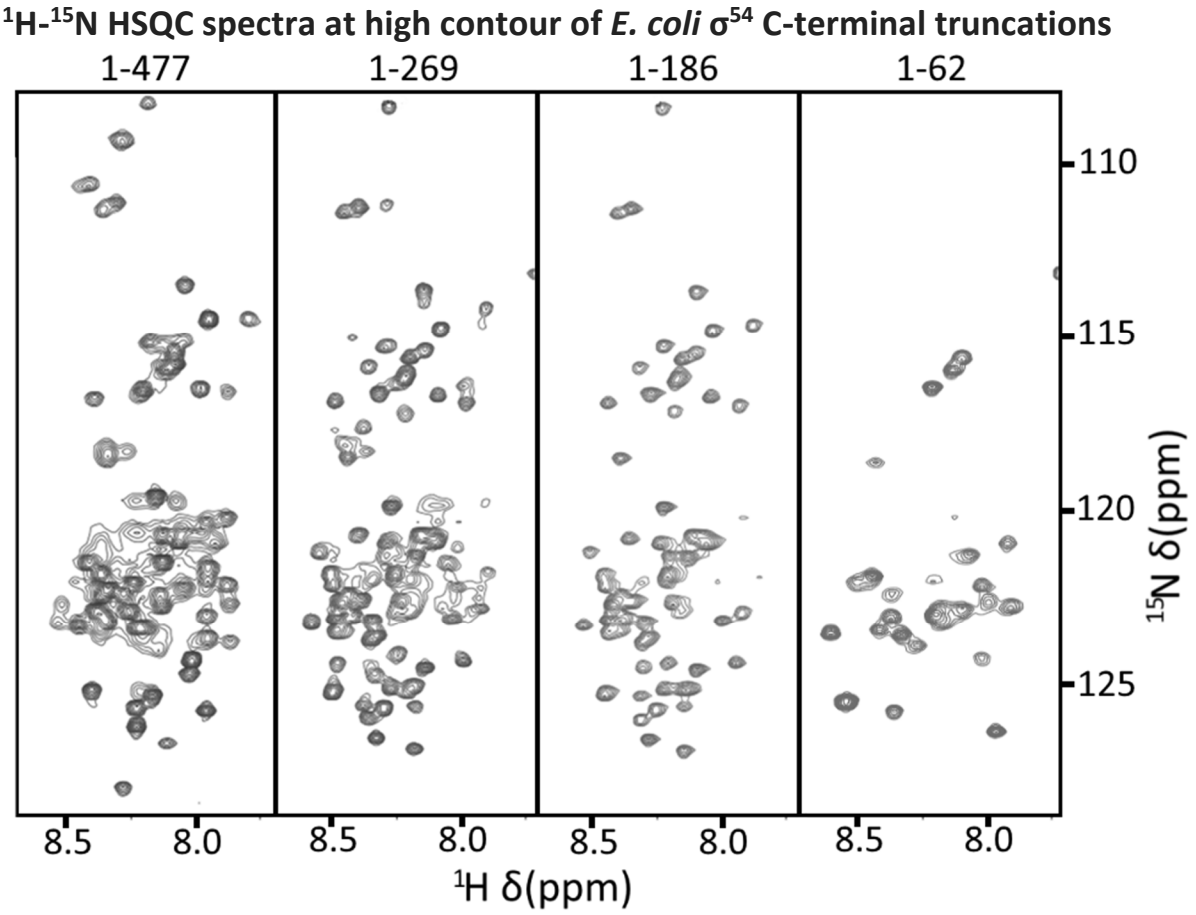


Figure 2.5 ^{1}H-^{15}N HSQCs of σ^{54} constructs each including the activator interacting domain and a different length of C-terminal domains. Spectra are shown at high contour levels where only the sharpest peaks of the flexible activator interacting domain remain visible.

2.3.2 The Activator Interacting Domain binds core RNA Polymerase but only when part of full length σ^{54}

To determine whether the activator interacting domain makes contacts in the context of the full RNA polymerase (RNAP), we collected ^{1}H-^{15}N HSQC spectra of *E.c.* σ^{54}(1-269), which includes the full AID and core binding domain, in the presence of increasing concentrations of *E.c.* RNA polymerase. In the absence of RNAP, the chemical shifts of peaks from the folded CBD align well with those of the CBD alone, with the peaks showing low shift dispersion contributed by the AID (Figure 2.4). With increasing RNAP concentrations, the peaks from the folded CBD decrease, and vanish at a 1:1 ratio of RNAP:σ^{54} (Figure 2.6). This is expected because the 400 kDa polymerase tumbles slowly, giving rise to severe line broadening of immobilized residues. However, the low dispersion peaks associated with the AID remain relatively sharp even with excess RNAP present. These peaks remain poorly dispersed making the number of

residues with remaining resonances difficult to determine, but the number roughly matches with the number from the AID alone σ^{54}(1-62).

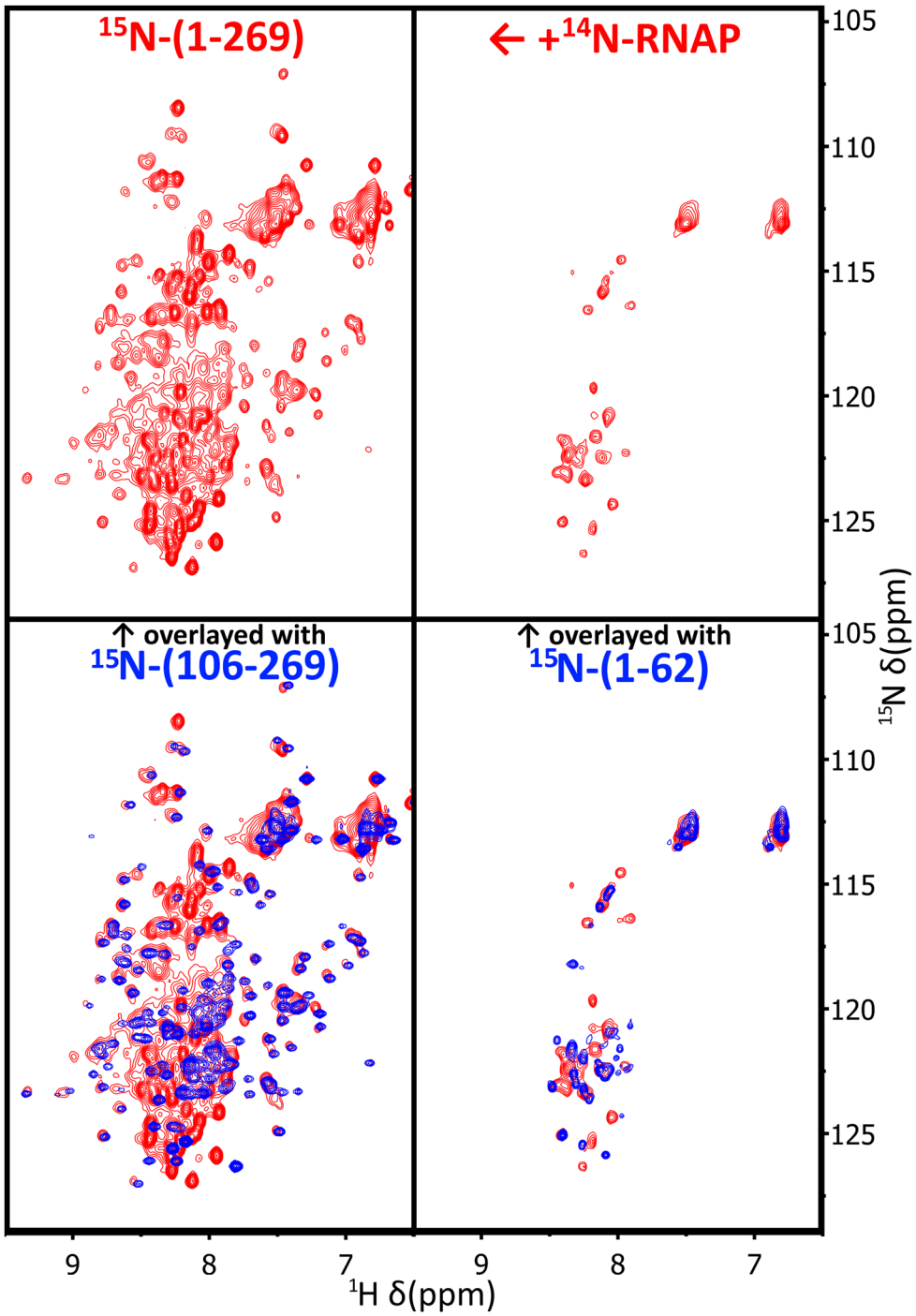


Figure 2.6 ^{1}H-^{15}N HSQC of *E.c.* σ^{54}(1-269) construct containing the activator interacting domain and core binding domain (top left, red). Most well-dispersed peaks broaden in the presence of *E.c.* RNAP (top right, red). These broadened peaks align with peaks from the core binding domain alone (bottom left, blue) and the poorly dispersed peaks that remain roughly align with peaks from the activator interacting domain (bottom right, blue).

To explore interactions between the AID and RNAP that require the presence of the remaining σ^{54} domains, as suggested by the crystal structure (Yang et al. 2015), we performed similar experiments with full length ^{15}N-labeled *E.c.* σ^{54} and RNA polymerase. With excess RNAP present some but not all of the peaks corresponding to the AID and linker are broadened (Figure 2.7). The ≈ 10 remaining peaks, many of which are broadened by the presence of the activator NtrC1^{C} and ADP-BeF$_3^-$, likely correspond to the N-terminal residues before the start of, and possibly including, the N-terminal helix of the AID. Thus the presence of the CBD alone is not sufficient to promote AID binding to RNAP, consistent with the suggested AID contacts with the -12 DNA binding domain in the holoenzyme.

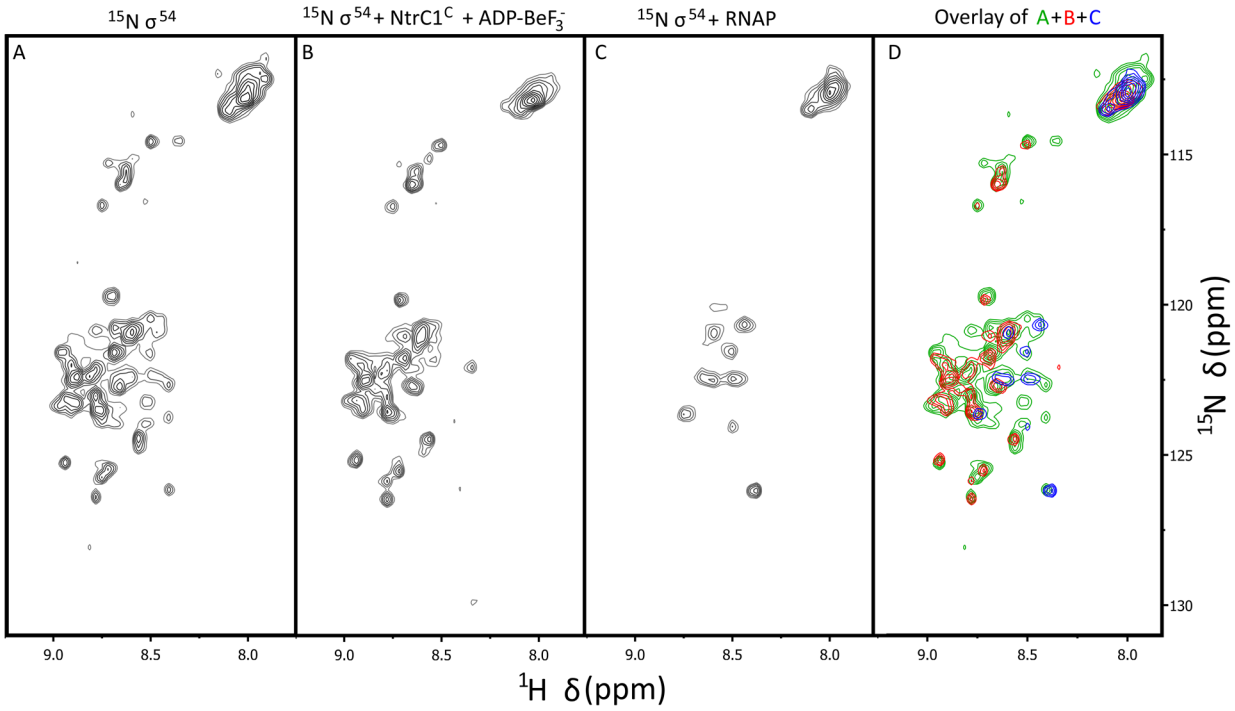


Figure 2.7 ^{1}H-^{15}N HSQCs at high contour levels showing the *E.c.* σ^{54}(1-477) construct with most well-dispersed peaks corresponding to folded domains below the cutoff (A). These peaks correspond to the AID and linker and are almost all gone in the *E.c.* $\sigma^{54}\Delta$AID(106-477) spectrum (not shown). Some of the σ^{54} AID peaks broaden in the presence of NtrC1^{C} and ADP-BeF$_3^-$ (B) and RNA polymerase (C) indicating that the AID binds to NtrC1^{C} in the ATP state and the core RNA polymerase as part of the *E.c.* σ^{54}-RNA polymerase complex. An overlay (D) of the spectra of free σ^{54} (green), σ^{54} bound to NtrC1^{C} and ADP-BeF$_3^-$ (red), and σ^{54} bound to RNAP (blue) suggests that the activator and RNAP may interact with different regions of the AID, presumably the N-terminal and C-terminal helices respectively.

To test whether the AID-RNAP interaction could occur in trans, NMR spectra of a truncated AID construct from *E.c.*, σ^{54}(11-48), were collected in the presence of polymerase and a σ^{54} N-terminal deletion that removed the AID. Alone, peaks from ^{15}N σ^{54}(11-48) show the expected low dispersion in the hydrogen dimension of the ^{1}H-^{15}N

HSQC spectrum (Figure 2.8). The peaks neither shift nor broaden in the presence of excess σ^{54}(106-477), showing there is no interaction between the AID and the folded domains of σ^{54} in the absence of core RNA polymerase. When excess RNA polymerase is added, there are only very minor perturbations to the chemical shifts and slight line-broadening. The slight broadening in this spectrum affects all peaks, and can be attributed to the increased viscosity of the samples containing high concentrations of RNA polymerase rather than any interaction between AID and the high molecular weight RNA polymerase holoenzyme. The ordering of the AID relative to the holoenzyme is greatly diminished when it is present as a separate molecule, indicating a weak interaction that depends on the AID being tethered near the holoenzyme by the linker region for interaction.

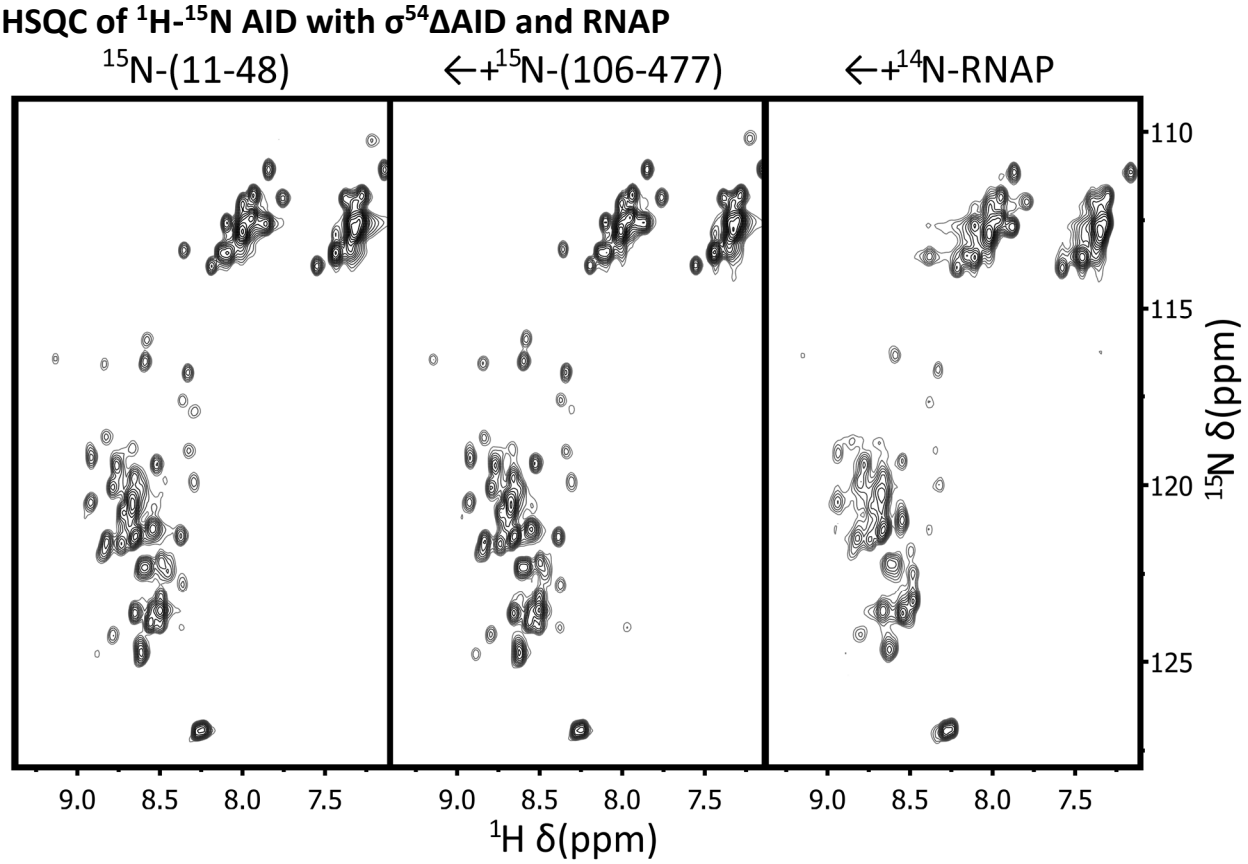


Figure 2.8 ^{1}H-^{15}N HSQCs at high contour levels showing that the *E.c.* σ^{54}(AID) construct (left) does not bind to *E.c.* σ^{54}ΔAID (middle) even in the presence of excess *E.c.* RNA polymerase (right). Broadening in the spectrum in the presence of RNAP (right) is slightly broadened due to the high viscosity of the solution in the NMR tube.

2.3.3 A Segment of the Activator Interacting Domain Drives Complex Formation with NtrC1^{C} in its ATP state

To study the interaction between the activator interacting domain and the transcriptional activators, we analyzed ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled *A.a.* σ^{54} constructs containing the AID in the presence of unlabeled central ATPase domain from the transcriptional activator NtrC1. The activator was maintained in its ATP state by the addition of the non-hydrolyzable ATP analog ADP-BeF_3^-, which causes NtrC1^{C} to assemble into heptamers with uniform ATP-like sites (Chen et al. 2007) that mimic the binding behavior of full length hexamers. Alternately, NtrC1^{C} forms an analogous, uniform ATP state heptamer when a single amino acid mutation of glutamate 239 to alanine (E239A) is present. This variant binds to, but does not hydrolyze, ATP and has biochemical properties and a structure resembling WT NtrC1^{C} with ADP-BeF_3^- (Chen et al. 2010). Titrating either heptameric NtrC1^{C} into a solution of the ^{15}N labeled AID caused changes in the chemical shifts and linewidths of some AID peaks. Broadening arises from slowed tumbling of the small AID when bound to the higher molecular weight ($\approx$ 210 kDa) NtrC1^{C} heptamer. With *A.a.* AID(1-62) many peaks in the σ^{54} construct remain sharp and unshifted even at a ratio of 1 AID per heptamer ring. This indicates that many residues in the AID do not contact NtrC1^{C}, and hence remain flexible (giving sharp resonances) even in the complex. To identify the key residues required to form the σ^{54}-activator encounter complex, we made constructs reducing the number of residues from the AID, while verifying that these AID constructs still bound to the activator ATPase oligomers.

2.3.4 Minimal AID construct

We also considered available data to try to identify what region of the σ^{54} activator interacting domain might be responsible for specific binding to activators. Although mutagenesis failed to identify individual residues critical for binding, changes that have a moderate effect occur mostly in the range of residues 20-40. Furthermore, this region contains two predicted helices (Buchan et al. 2013)(Jones 1999). We therefore prepared a construct of *A.a.* σ^{54} that begins with the first predicted α helix and ends after the second one, comprising residues 16-41, σ^{54}(16-41). As shown below, this construct still binds well to the activator. The ^{1}H-^{15}N HSQC peaks for this construct cluster in the same region of the spectrum as the longer versions, consistent with a lack of folded structure as in the full AID (Figure 2.9).

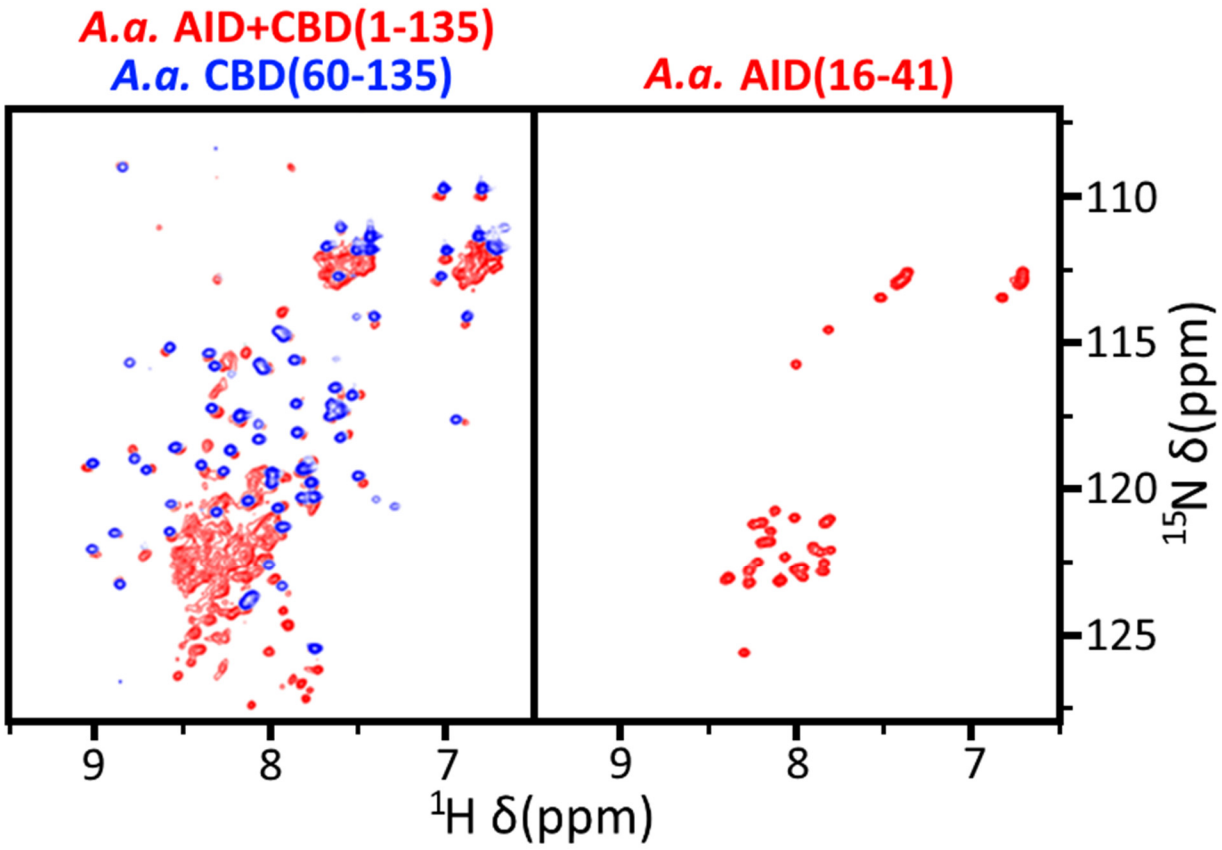


Figure 2.9 Comparison of the σ^{54}(AID) ^{1}H-^{15}N HSQC (right) to an overlay of the σ^{54}(AID) and σ^{54}(AID+CBD) ^{1}H-^{15}N HSQCs (left) showing that the AID can account for the difference between these two spectra.

For the peptide alone the ^{1}H-^{15}N HSQC peaks from all 25 of these residues are sharp and distinct. We assigned amide peaks in σ^{54}(16-41) (Figure 2.10) by using sequential connectivities in 3D ^{15}N-resolved [^{1}H-^{1}H]-NOESY spectra together with residue type identifications from intra-residue peaks. Neither patterns of NOEs or chemical shifts indicate any regions of substantial secondary structure for the peptide alone in spite of the two predicted helices.

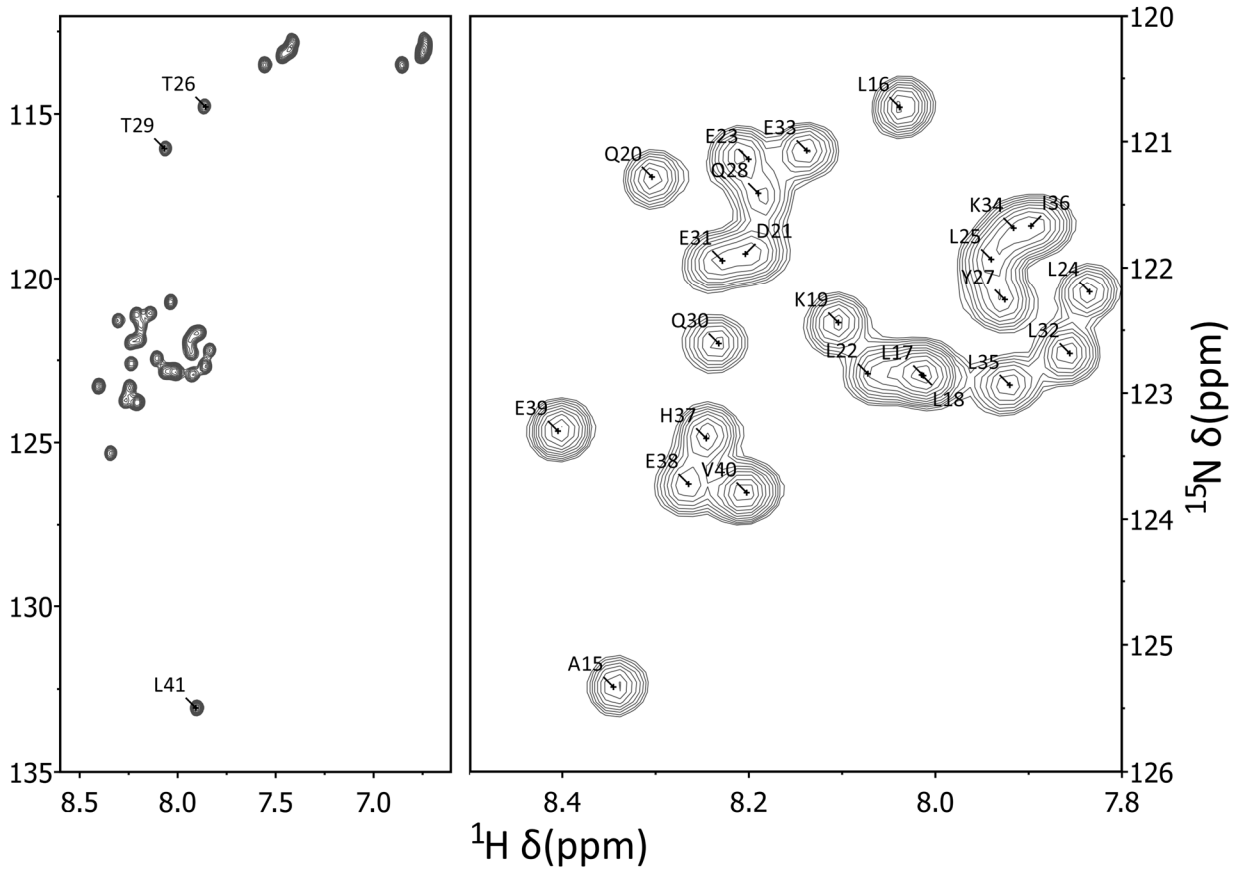


Figure 2.10 Amide assignments of the AID(16-41) construct obtained from 3D ^{15}N-resolved [^{1}H-^{1}H]-NOESY data. A tall spectrum (left) shows the assignments of the outlying peaks for T26, T29 and the C-terminal residue L41. A zoomed in spectrum (right) shows the assignments of the remaining peaks. The identity of some of the leucines, in particular L17, L18, and L32, could not be assigned with absolute certainty.

2.3.5 Characterization of the σ^{54}-NtrC1^{C} Binding

In the presence of transcriptional activator NtrC1^{C} locked in its ATP state with the analog ADP-BeF_3^-, the 3.5 kDa σ^{54} AID (16-41) peptide elutes from a size exclusion column in the same fraction as the activator oligomers, but elutes much later than NtrC1^{C} when the activator is in the apo or ADP state. Thus a complex between NtrC1^{C} and σ^{54} forms, even with only 25 residues from the predicted helices in the activator interacting domain.

We also cloned a version of the reduced AID segment 16-41, adding an N-terminal tryptophan followed by a cysteine before the start of the σ^{54} sequence at residue 16. The cysteine was derivatized with the fluorescent marker Alexa Fluor 488 by forming a linkage between the cysteine and a maleimide on the dye. Binding was assayed by adding increasing concentrations of NtrC1^{C} (with ADP or ATP analog ADP-BeF_3^- at

500 μM) to 0.16 μM Alexa488-σ^{54}(16-41). Each sample was run on a native gel, and visualized using a Typhoon at 488nm. In the gel, when ADP-BeF_3^- is present, the $NtrC1^C$ oligomer band fluoresces at 488nm showing that the Alexa488 labeled σ^{54}(16-41) bound to $NtrC1^C$ as the complex traveled through the gel (Figure 2.11C). However, when only ADP is present, the $NtrC1^C$ oligomer has little emission at 488nm, indicating binding in the ADP state is substantially reduced relative to the ATP state, as expected from previous work (Chen et al. 2007)(Chen et al. 2010).

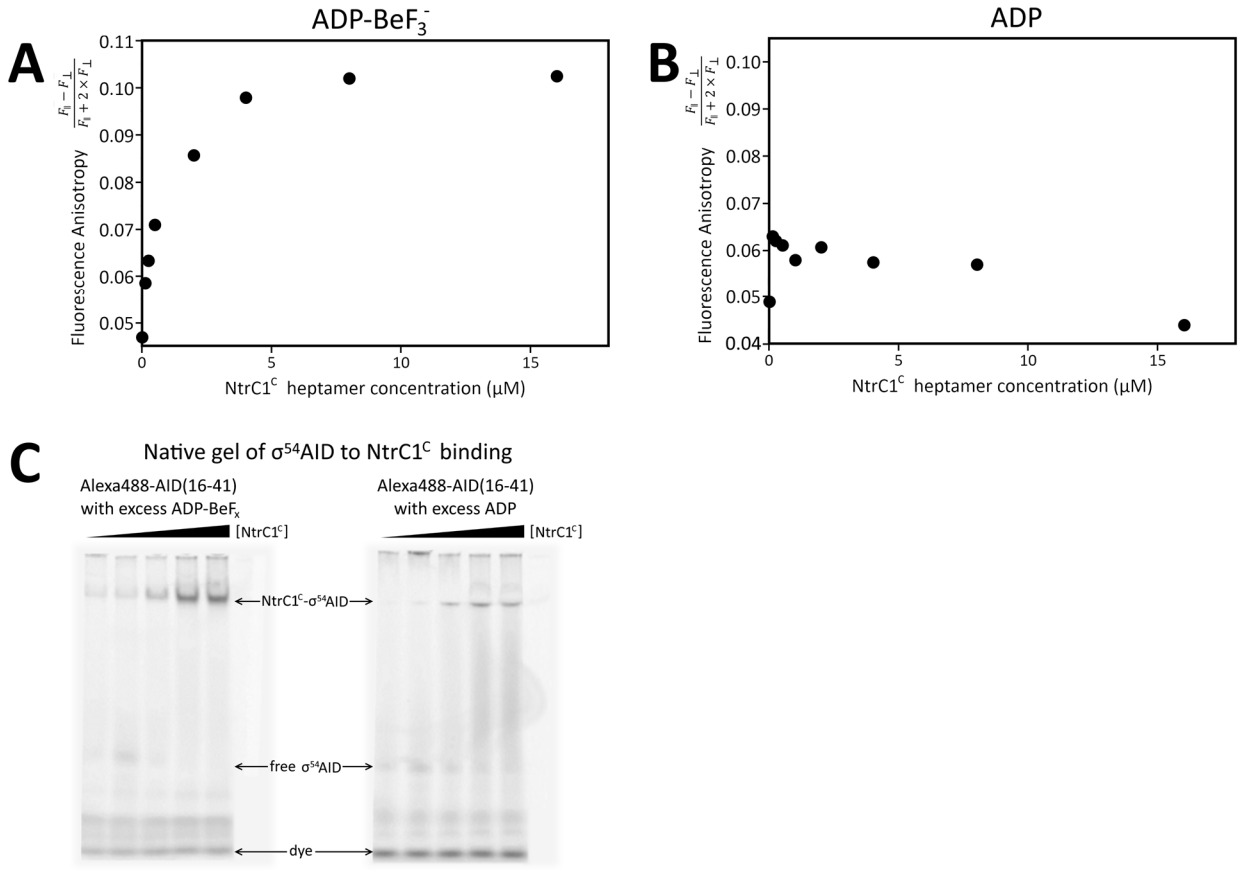


Figure 2.11 Fluorescence anisotropy data for 0.16 μM Alexa488-labeled AID(16-41) with (A) 500 μM ADP-BeF_3^- or (B) 500 μM ADP and increasing concentrations of $NtrC1^C$ heptamer. The K_d of the binding is the concentration where the fluorescence anisotropy is halfway between its maximum value with high excess $NtrC1^C$ (fully bound AID(16-41)) and its minimum value with no $NtrC1^C$ (fully unbound AID(16-41)), which occurs at 1 μM in the presence of ADP-BeF_3^-. No binding is detected in the presence of ADP only. (C) Native gel of 5 μM AID(16-41), 500 μM ADP-BeF_3^- or ADP, and increasing concentrations of heptameric $NtrC1^C$: 0.5, 1, 3.5, 38.5, and 50 μM. The sixth lane was loaded with only $NtrC1^C$ and no Alexa488-AID(16-41). The gel was visualized at 488nm on a Typhoon fluorescence scanner.

To assess the binding affinity of our reduced AID construct, we measured the changes in the fluorescence anisotropy of Alexa488 labeled σ^{54}(16-41) upon adding $NtrC1^C$ in different nucleotide states. Binding of the small, fluorescently labeled σ^{54}(16-41) peptide

to the high molecular weight NtrC1^{C} greatly lengthens the rotational time of the peptide, resulting in a change in anisotropy in the dye fluorescence. Titrations were done by adding increasing concentrations of ADP-BeF_3^--NtrC1^{C}, ADP-NtrC1^{C} or just NtrC1^{C} to a fixed concentration of peptide, and following the anisotropy change. In the presence of ADP-BeF_3^-, σ^{54} AID(16-41) binds with a K_d of 1 μM (Figure 2.11A). In the absence of any nucleotide, or in the presence of ADP, we observed no change in the fluorescence anisotropy, and therefore no binding, even with a significant excess of NtrC1^{C} (Figure 2.11B). This confirms that the minimal AID construct σ^{54}(16-41) is sufficient to form the activator complex.

Fluorescence anisotropy experiments were also performed with full length σ^{54} carrying a K2C mutation and with an Alexa Fluor 488 dye attached at residue 2. The anisotropy changes during titrations with ADP-BeF_3^--NtrC1^{C} indicate a similar K_d of 2 μM (Figure 2.12). The equivalence of the binding constants for σ^{54}(16-41) and full length σ^{54} to the 'ATP' form of NtrC1^{C} shows that residues between 16 and 41 are responsible for binding to the activator.

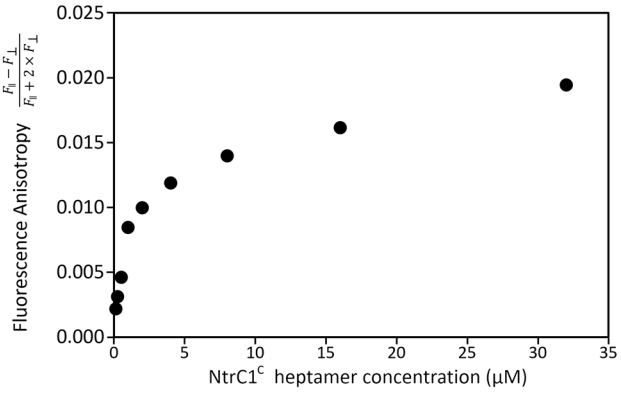


Figure 2.12 Fluorescence anisotropy data for 5 μM Alexa488-labeled full length σ^{54}(K2C) with 500 μM ADP-BeF_3^- (left) and increasing concentrations of NtrC1^{C} heptamer. The K_d of the binding was calculated using non-linear least squares regression from the observed fraction bound between its maximum value with high excess NtrC1^{C} (fully bound σ^{54}) and its minimum value with no NtrC1^{C} (fully unbound σ^{54}). The K_d is 2 μM in the presence of ADP-BeF_3^-.

2.3.6 NMR characterization of the AID-NtrC1^{C} complex

Given the binding affinity of the peptide for NtrC1^{C} and the high molecular weight of the complex, we thought it might be possible to use exchange transferred NOEs within the peptide to characterize the bound state. However, titrating in sub-stoichiometric NtrC1^{C} with ADP-BeF_3^- produced no new NOE peaks even at longer mixing times, indicating slow exchange between the free AID and the AID-NtrC1^{C} encounter complex.

We also examined HSQC spectra collected with varying ratios of peptide and activator. Addition of $NtrC1^C$ with ADP or $NtrC1^C$ alone had no effect on the peptide spectrum (consistent with fluorescence data), but the addition of ADP-BeF_3^- together with $NtrC1^C$ (not shown) or ATP with $NtrC1^C$(E239A) (Figure 2.13), reduced the intensity of specific peaks. Most peaks from the ^{15}N-labeled σ^{54}(16-41) are so broad that they cannot be observed when excess ATP and $NtrC1^C$(E239A) are present, but amides for residues E38, E39, V40, and L41, remain sharp even in the complex and amides for K34, L35, I36 and H37 appear but are very weak. The peaks with the least broadening are from the C-terminal part of the σ^{54}(16-41) AID construct, indicating that the N-terminus and the first predicted helix are likely the main contributors to binding affinity.

To try to directly characterize the high molecular weight activator-σ^{54} AID complex we made ^{2}H/^{15}N σ^{54}(16-41) with specific ^{1}H-^{13}C methyl labels by expressing it in ^2H_2O/$^{15}NH_4Cl$ in the presence of labeled metabolic precursor α-ketoisovalerate to label the protein only at the δ-methyl of leucine and γ-methyl of valine (Goto et al. 1999). The labeling approach produces a deuterated σ^{54}(16-41) peptide with nine leucines and one valine each labeled with ^{1}H and ^{13}C at the δ_1 or δ_2 methyls of leucine and γ_1 or γ_2 methyls of valine. We then ran ^{1}H-^{13}C methyl-TROSY HMQC experiments (Tugarinov et al. 2003) to detect signals from the labeled residues in the complex.

For the peptide alone the leucine δ methyl peaks were poorly dispersed and all located close to the values predicted for random coil chemical shifts (Figure 2.14), as might be expected given the ^{1}H-^{15}N HSQC spectrum. The single valine is easily identified from its γ_1 and γ_2 peaks that appear with equal intensities and with nearly equal ^{1}H chemical shift but different ^{13}C chemical shifts. The nine sets of leucine peaks overlap significantly and cannot clearly be distinguished, though the nine δ_1 peaks are distinct from the nine δ_2 peaks and are consistent with chemical shifts expected for a leucine in a random coil. Though individual leucine δ_1 and δ_2 peaks cannot be identified, integration of the region of overlapping peaks suggests that they do correspond to 9 times as many methyls as a single valine peak, which is expected from the ratio of nine leucines and one valine in the σ^{54}(16-41) sequence.

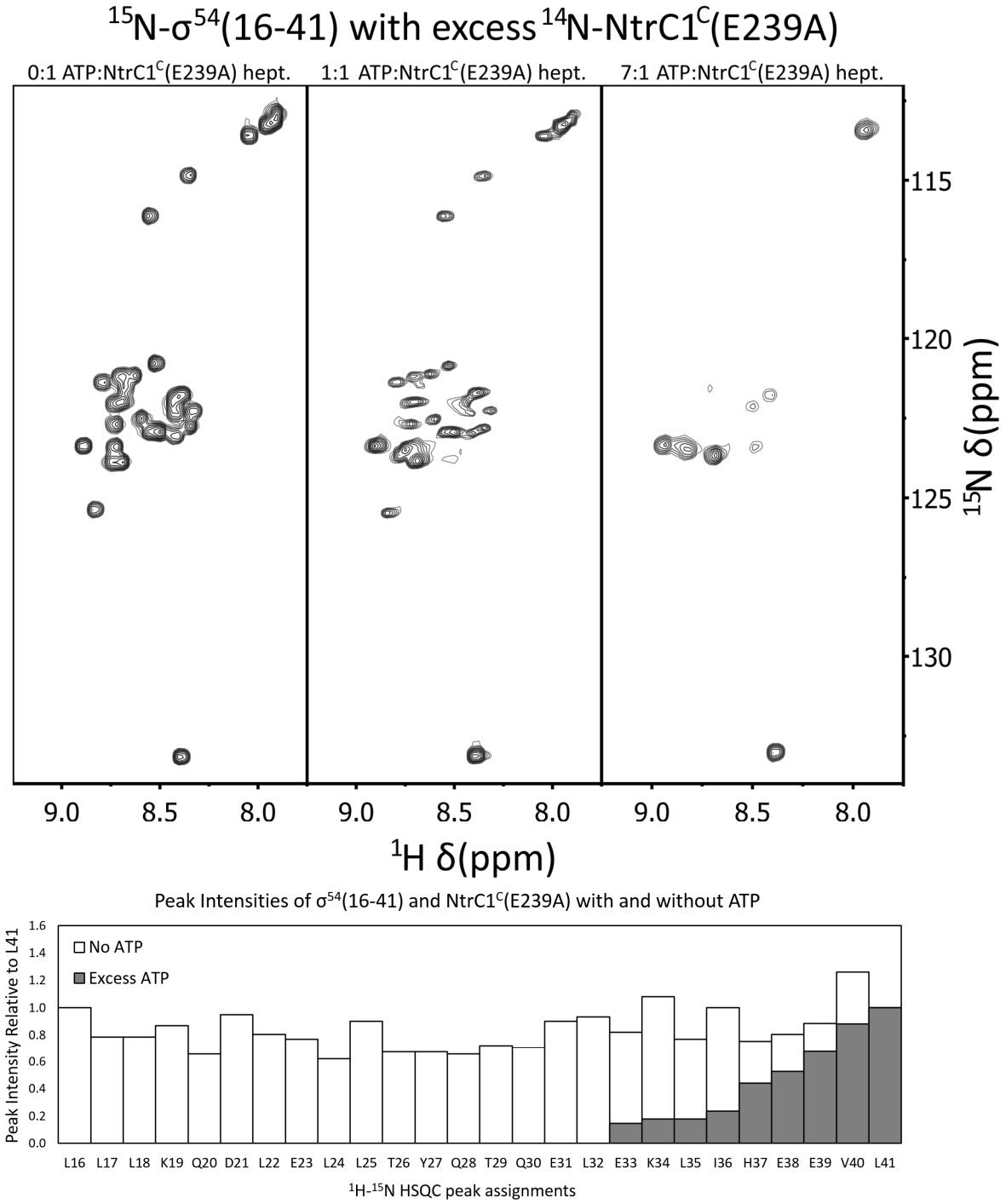


Figure 2.13 ^{1}H-^{15}N HSQC of *A.a.* ^{15}N-σ^{54}AID(16-41) and ^{14}N-NtrC1^{C}(E239A) heptamer without nucleotide (left) with 1:1 ATP:NtrC1^{C}(E239A) heptamer (middle) and with 7:1 ATP:NtrC1^{C}(E239A) heptamer (right) showing the disappearance of most AID peaks when bound to NtrC1^{C}(E239A) trapped in the ATP state. The remaining AID peaks at high ATP concentrations are localized to the second predicted helix and correspond to E38, E39, V40, and L41, with broad peaks present for K34, L35 and I36. One remaining glutamine side chain peak likely corresponds to Q30. Peak intensities of each residue ^{15}N-σ^{54}AID(16-41) with ^{14}N-NtrC1^{C}(E239A) in the presence (gray) and absence (white) of ATP are shown below.

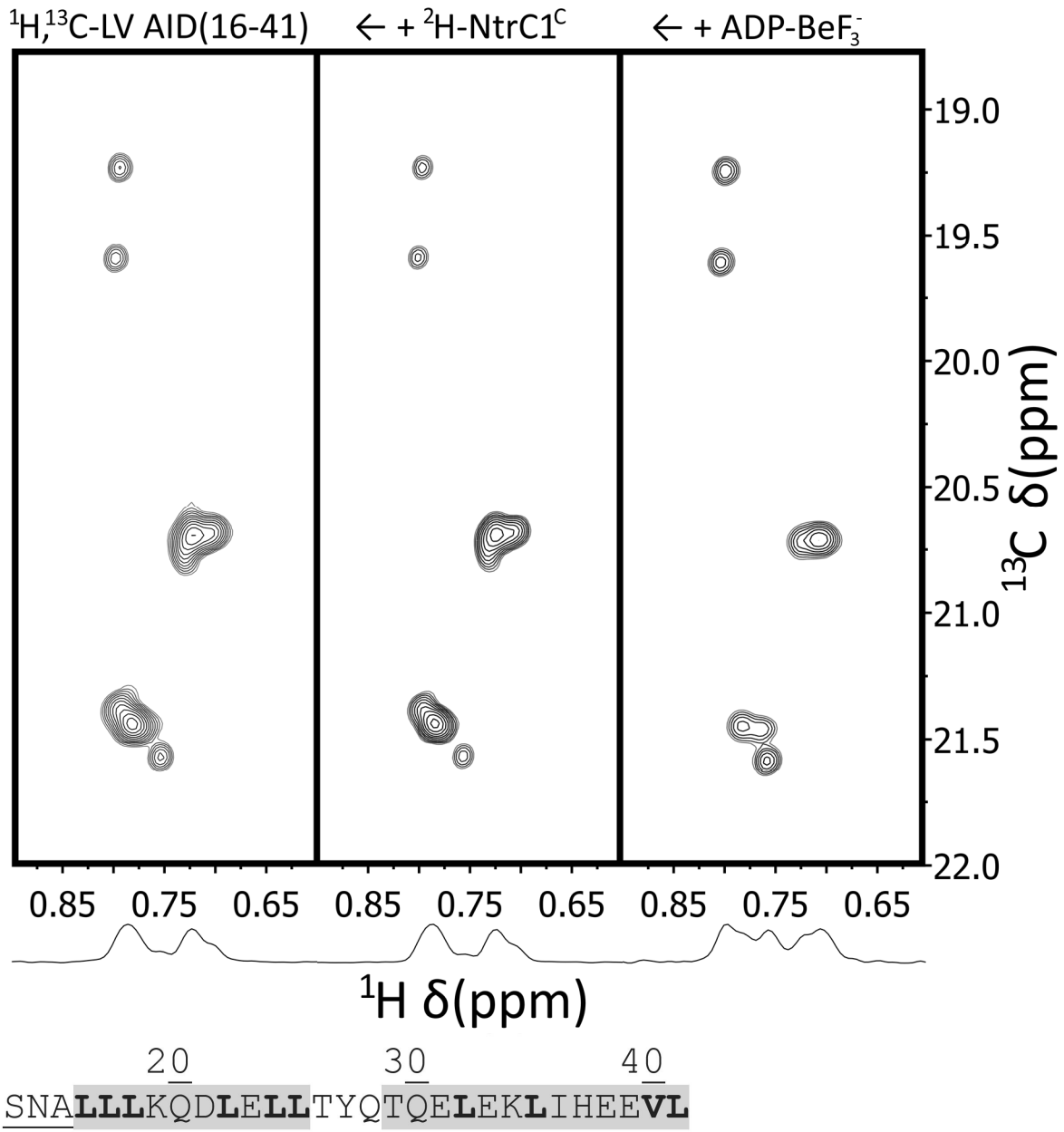


Figure 2.14 2D ^{1}H-^{13}C methyl-TROSY HMQC spectrum of AID(16-41) with ^{1}H-^{13}C labeling of δ_1 or δ_2 methyls of leucine and γ_1 or γ_2 methyls of valine and ^{2}H and ^{12}C labeling of all other carbons and hydrogens. Integration of all leucine and valine side chain methyls matches the expected number of nine leucines per one valine in the AID(16-41) alone (left). The spectrum is unaffected by addition of excess ^{2}H-NtrC1^{C} (middle) but integration indicates six leucines are broadened by the addition of ^{2}H-NtrC1^{C} and ADP-BeF$_3^-$ (right) leaving three leucines and one valine. The sequence of the AID peptide with its two predicted α-helices highlighted in gray is shown below. The nine leucines and one valine are in bold. The three underlined residues "SNA" in front of the first helix are the result of TEV cleavage and are not part of the true *A.a.* σ^{54} sequence.

Addition of NtrC1^{C} alone had no effect on the chemical shifts of the leucines and valines in the labeled σ^{54}(16-41) minimal AID peptide, again indicating no or very weak interaction between the two in the absence of ATP (Figure 2.14). In the presence of the ATP analog ADP-BeF_3^- and ^{2}H-NtrC1^{C}, the 2D ^{1}H-^{13}C methyl-TROSY HMQC spectrum changes, reflected also in the 1D projections. Many peaks in the leucine δ_1 and δ_2 regions are broadened, but three remain relatively sharp along with the γ_1 and γ_2 peaks from the single valine. The integration of the leucine δ_1 and δ_2 regions is consistent with the broadening (loss of signal) of six leucines. The three relatively non-overlapping leucine peaks show only small chemical shift changes. These observations are consistent with the ^{1}H-^{15}N HSQC data showing that NtrC1^{C} binding is localized to the N-terminus of the AID peptide, which has 6 leucines, and that the C-terminal tail of the AID, which has 3 leucines and 1 valine, remains free in the high molecular weight complex. Surprisingly, the six N-terminal leucines bound to the $\approx$ 210 kDa NtrC1^{C} complex could not be detected even using TROSY experiments designed to minimize the broadening of high molecular weight species. This suggests that in spite of the relatively binding there may be conformational exchange in the complex such that contact residues are broadened.

2.4 Discussion

Previous structural, biochemical, and genetic studies of σ^{54} have identified functional regions in the protein that play different roles in transcription: interaction with activators; binding to core polymerase; DNA interactions in the –12 region of the promoter where opening occurs; and sequence specific recognition of the –24 region of the promoter. The latter three functions are carried out by folded domains of σ^{54}; structures of the core binding domain (Hong, Doucleff, and Wemmer 2009) and –24 recognition domain (Doucleff et al. 2007) have been solved as independent domains. DNA opening is primarily controlled by the –12 DNA binding domain of σ^{54} in the RNAP holoenzyme (Wong, Gralla, and Tintut 1994). The activation event that brings about these conformational changes is initiated through the first 50 N-terminal residues of σ^{54}, the activator interacting domain (AID), and a transcriptional activator. The AID has relatively low sequence conservation, and the linker region between the AID and core binding domain is quite variable in length. Neither genetic studies nor a recent low resolution crystal structure of the polymerase-σ^{54} complex (Yang et al. 2015) has provided insight into how the AID functions. Here we address the nature of the AID and its interaction with the activator ATPase domain. These studies of the σ^{54} AID are

a step toward understanding the changes in the σ^{54}-polymerase holoenzyme that lead to transcription initiation.

2.4.1 The Activator Interacting Domain is Intrinsically Disordered in σ^{54} Alone

^{1}H-^{15}N HSQC spectra from all *A.a.* and *E.c.* σ^{54} constructs that include the N-terminal domain, ranging in size from full length down to a minimal construct of the AID, display many relatively sharp resonances with low chemical shift dispersion, particularly in the proton dimension. The low shift dispersion is a hallmark of an unfolded domain, the lack of secondary structure leading to a similar environment for the backbone amides (Wishart, Sykes, and Richards 1991). Rapid local motion of the backbone of unfolded domains also reduces the linewidths from those expected for a folded domain of the same molecular weight (Dyson and Wright 2004). By comparing spectra of σ^{54} constructs containing the AID with those lacking it, it is clear that the residues in the AID and linker domain of σ^{54} account for the vast majority of the poorly dispersed, unstructured residues. The peaks from the AID have the same chemical shifts when part of constructs with the σ^{54} CBD or the full length σ^{54} present. Addition of RNA polymerase to ^{15}N labeled σ^{54} and its interaction with the well-folded domains of σ^{54} leads to slower tumbling due to high molecular weight, and almost all of the resonances from these domains disappear, including some peaks from the AID. It has been proposed that such an interaction could lead to the AID blocking the DNA from accessing the active site of the RNA polymerase and inhibiting transcription initiation (Yang et al. 2015). However, the AID does not bind the holoenzyme when present *in trans* and therefore may only have a low affinity for it. In this case, in the holoenzyme the AID would remain disordered and undocked some of the time allowing it to bind the transcriptional activators.

^{1}H-^{15}N HSQC spectra of ^{15}N-labeled *A.a.* σ^{54} constructs were also studied in the presence of the ATPase domain from *A.a.* activator NtrC1^{C}. In this case, the AID peaks are strongly broadened due to immobilization in the presence of NtrC1^{C} in its ATP state (mimicked by ADP-BeF_3^-), but they are unaffected by apo NtrC1^{C} or NtrC1^{C} in the ADP state where the activator-σ^{54} interaction is weak. The observed broadening indicates that the AID becomes immobilized in the complex with the oligomeric, high molecular weight transcriptional activator oligomer, but only when the activators are in their ATP state with the GAFTGA loops raised above the ring around the central pore (Lee et al. 2003)(Chen et al. 2007).

The intrinsically unfolded domain becoming immobilized and partly structured upon encountering a specific binding partner is common behavior for this class of domains (Dunker et al. 2002). One common role for intrinsically unfolded domains is to facilitate assembly of functional complexes (Wright and Dyson 2015), for example to initiate transcription in eukaryotes. In that case, transcription factor activation domains serve to form tethers with other proteins required to assemble a polymerase complex competent for initiation (Liu et al. 2006)(Tantos, Han, and Tompa 2012). For bacterial σ^{54}-polymerase, it is expected that the polymerase holoenzyme is normally bound at the promoter waiting for the action of the activator, which may already be bound to the enhancer region in its unactivated dimeric state. Upon receiving its signal, the activators assemble to the active hexamer (Porter et al. 1993), which should have primarily ATP bound since ATP is much more abundant than ADP in the cell. At this point, the flexibility of the AID of σ^{54} would facilitate making the initial 'encounter complex' with the ATPase domain of the assembled activator, completing assembly of all the components needed for initiation of transcription upon subsequent ATP hydrolysis. At present it is not clear whether there are contacts between the activator ATPase domain and other parts of the polymerase holoenzyme, how much ATP hydrolysis is required for promoter opening, and the extent to which the complex changes structure when ATP hydrolysis occurs. Studies of the activator ATPase domains have shown that the GAFTGA loops that contact the AID in the initial complex are raised above the ring in their ATP bound state (Chen et al. 2007) but move down when ADP is bound (Lee et al. 2003). A recent structure of the $NtrC1^{C}$ ATPase crystallized in the presence of ADP-BeF_3^- gave a mixed nucleotide state for the hexameric ring, with a 'lock-washer' spiral structure (Sysoeva et al. 2013) reminiscent of the Rho-ATPase (Skordalakes and Berger 2006). For Rho a functional model has been developed in which sequential hydrolysis drives loops down pushing RNA through the central pore of the ring through multiple rounds of ATP hydrolysis. Although the organization of the ATPase domains is different, there is some structural similarity between these systems.

2.4.2 A minimal Activator Interacting Domain is sufficient for formation of the encounter complex

Mutational studies of the N-terminal activator interacting domain of σ^{54} show that no single residue is critical for activator binding and function, rather changes in numerous residues over a segment of about 25 amino acids result in modest decreases in activity (Xiao et al. 2009)(Syed and Gralla 1998)(Gallegos and Buck 2000)(Hsieh, Tintut, and

Gralla 1994). Secondary structure predictions using PSIPRED indicate that there may be two conserved helices in σ^{54}, predicted consistently across multiple species (Buchan et al. 2013). Our experiments with σ^{54} AID constructs show that just having residues from these two predicted helices is sufficient for activator binding in the presence of an ATP analog, but the NMR data do not suggest significant helix formation in the peptide alone. Just as with full length σ^{54}, the formation of the encounter complex between the minimal σ^{54} AID construct and the transcriptional activators requires activator subunits in the ATP state in which GAFTGA loops are extended above the ATPase ring. The dependence of peptide binding on ATP, as well as the similar binding affinities of the peptide and the full length σ^{54} to the transcriptional activators, confirms that we are observing native-like interactions of this segment of the AID rather than a non-specific binding event.

Examining the sequences of the *E.c.* and *A.a.* AIDs shows that if they fold into the predicted α-helices they would be amphipathic. Regular repeats of hydrophobic residues (primarily leucine) and hydrophilic residues (primarily glutamine) cluster on opposite sides of the predicted α-helices (Figure 2.15), which would be connected by a short linker. This suggests that formation of the encounter complex between σ^{54} and the GAFTGA loops of the transcriptional activators could be driven by hydrophobic interactions between the leucine-rich side of the AID and the hydrophobic GAFTGA loops that form a surface around the pore in the ATPase ring.

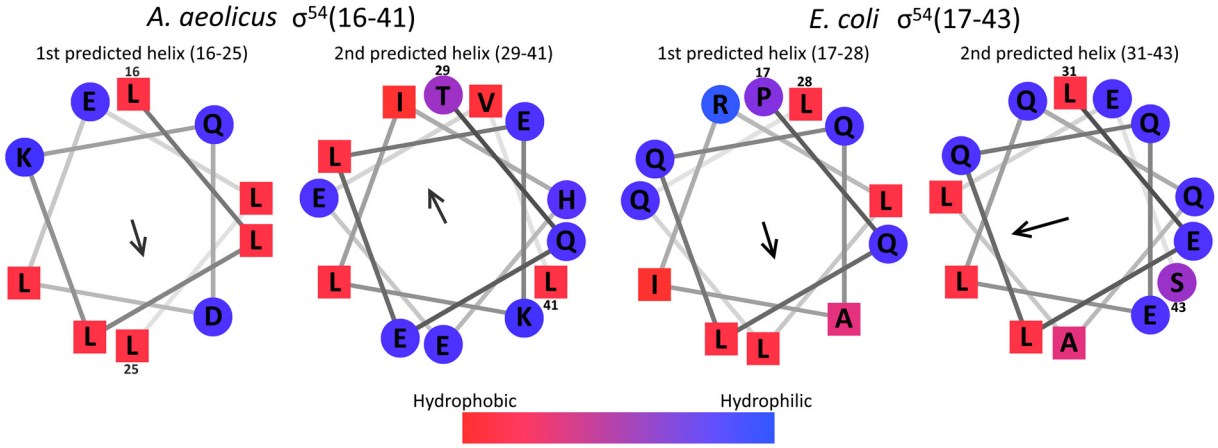


Figure 2.15 Helical wheel diagrams of the two PSIPRED predicted helices of the activator interacting domain in *A. aeolicus* (left) and *E. coli* (right). Residues are colored by their Kyte-Doolittle hydrophobicity score (more red, more hydrophobic; more blue, more hydrophilic) with hydrophobic residues shown as squares and hydrophilic residues shown as circles (Kyte and Doolittle 1982). Amphipathic nature of the predicted helices is indicated by the arrows pointing towards their hydrophobic side.

2.4.3 σ^{54}(AID) binding affinity and slow exchange

Full length σ^{54} and σ^{54}(16-41) bind NtrC1^{C} with approximately the same dissociation constant of ≈ 1-2 μM, while neither binds to the activator with measurable affinity in the ADP or apo state. This indicates that residues between 16 and 41 are not only sufficient for binding, but that they account for essentially the full binding affinity of σ^{54}. Thus it is likely that this segment of σ^{54} contains the large majority, if not the entirety, of the NtrC1^{C} binding surface during the formation of the initial encounter complex.

When the ^{15}N-labeled σ^{54} AID is present in excess of the ATP-state NtrC1^{C} rings, there are no discernible chemical shift changes in the unbound AID peaks. In these samples, σ^{54} AID is present in two states: the bound encounter complex and the unbound free form. The lack of chemical shift changes or broadening during the titrations shows that the AID is in slow exchange on the chemical shift timescale. This is somewhat surprising given the binding affinity of the peptide, if binding were even near the diffusion limit then the predicted dissociation rate would be sufficient to give intermediate to fast exchange behavior. To be consistent with the observed dissociation constant and the slow exchange behavior both the rate of binding and the rate of dissociation must be quite low. In the presence of excess NtrC1^{C}-ADP-BeF_3^- (above one ATPase ring per peptide) all but a few C-terminal peaks of the AID broaden and disappear, indicating complete binding of the AID.

2.4.4 Characterization of the σ^{54}-NtrC1^{C} Encounter Complex

The minimal AID construct σ^{54}(16-41) yielded NMR spectra with little overlap of amide residues and uniformly sharp peaks, which enabled doing peak assignments using conventional sequential methods (Wüthrich 1986). In the NOESY spectra we observed no crosspeaks between i and i+3 residues as would be expected in an α-helical structure, in which the α-proton of a residue is in proximity of the amide of the residue 3 amino acids later in the sequence. This indicates that the free AID peptide is not significantly populating folded α-helices by itself in solution, but this does not rule out the formation of well-folded α-helix in the encounter complex with the ATPase.

In the presence of NtrC1^{C} in the ATP state, most of the σ^{54}(16-41) peaks were broadened upon binding to the high molecular weight NtrC1^{C} oligomers. However, a few peaks remained sharp enough for detection and underwent slight chemical shift changes. The relatively sharp resonances indicate that these residues remain quite flexible even

when the AID is in the encounter complex. Our assignments (Figure 2.10) show that these residues arise from E38, E39, V40, and L41 at the end of the second predicted helix. Further very weak (broadened) resonances are consistent with chemical shifts of K34, L35, I36 and H37. To try to directly observe the bound state, methyl-TROSY experiments were done on uniformly deuterated σ^{54}(16-41) labeled with a ^{13}C-$^{1}H_3$ methyl at the δ carbon on the leucine and at the γ carbon of the valine side chains. In the free peptide these peaks exhibited low chemical shift dispersion as expected, consistent with other evidence that the AID is unstructured. Integration of the peaks shows that all nine leucines and one valine were labeled as expected. When AID binds to the transcriptional activators there are six leucine peaks that broaden and disappear from the methyl-TROSY spectra, while three leucine peaks and one valine remain sharp. There are six leucines in the first predicted helix and three leucines and one valine in the second predicted helix. One of the remaining leucines has a slightly different shift (unaffected by complex formation) and is very likely the C-terminal residue. The other two leucines that are observed, together with the only valine, which neighbors the C-terminal leucine, remain almost as sharp as in the free peptide. The combination of ^{15}N and ^{13}C observations show a gradient in mobility of the bound peptide residues, with flexibility starting around residue 33. The backbone appears to be affected more than the methyl-containing sidechains, but this might reflect sidechain dynamics.

In methyl-TROSY experiments, the methyl ^{1}H-^{13}C correlations are not broadened as much by the slow tumbling times of the high molecular weight activator-σ^{54} AID complex as those from other ^{1}H-^{13}C or amide ^{1}H-^{15}N pairs. Nevertheless, we still failed to detect even the methyl peaks corresponding to leucines in the N-terminal half of the AID in the ATPase encounter complex. This obviously prevents further analysis of the induced structure in the AID. It must also mean that the AID in the encounter complex is not bound to the transcriptional activator in a single, well-defined configuration, but instead is bound in multiple different conformations each experiencing a different chemical environment. The lack of a single, well-defined induced structure in the AID encounter complex is consistent with an activation mechanism that involves changes in the AID-ATPase complex rather than serving a role simply as a defined docking site. Given that many AAA+ ATPases thread substrates through a central pore to act on them (Tucker and Sallai 2007), we believe that a dynamic activation mechanism in which the AID is threaded through the central pore of the transcriptional activator by successive rounds of ATP hydrolysis and motion of the GAFTGA loops is a possible explanation for the properties of this initial encounter complex.

2.4.5 Future Directions

Studies of the interaction of the intrinsically disordered activator interacting domain bound with $NtrC1^{C}$ provide some hints to the way transcriptional activators drive σ^{54} transcription initiation. Further structural studies may be possible, if conditions for forming a better-defined complex can be identified, either using NMR techniques or crystallography. A high resolution structure of the complex could provide a more complete picture of the specific interactions between σ^{54} and the activators and may give some further evidence about the activation mechanism. However, even a more complete understanding of the structure of the static encounter complex may not be enough to identify the precise mechanism of activation. Experiments that probe the dynamics of the AAA+ ATPase transcriptional activators and σ^{54} as they go through multiple ATP hydrolysis cycles, as well as *in vivo* experiments to test various changes to σ^{54} that disrupt activation, could be valuable techniques to study the dynamics of the mechanism of σ^{54} transcription initiation.

2.5 Materials and Methods

2.5.1 Protein expression and purification

Aquifex aeolicus and *Escherichia coli* σ^{54} constructs were cloned from full length σ^{54} plasmids and placed into pET28a expression vectors. *E.c.* σ^{54}(1-477), *E.c.* σ^{54}(1-269), *E.c.* σ^{54}(1-186), *E.c.* σ^{54}(1-62), and *A.a.* σ^{54}(1-135) all contained a C-terminal His_6 tag. *E. coli* Rosetta cells with the plasmid were grown at 37°C in 1L of isotope-labeled M9 minimal media to an optical density at 600 nm of 0.6 and induced with 0.5 mM isopropyl thiogalactopyranoside then harvested after 8-12 hours. Extracts for *A.a.* constructs were heated at 75°C for 20 minutes, and both *E.c.* and *A.a.* constructs were purified on a NiNTA column. Some constructs went to inclusion bodies and were purified on a NiNTA column under denaturing conditions in 8M urea, then refolded on the column by changing to buffer without urea.

E.c. σ^{54}(11-48), *A.a.* σ^{54}(10-47), *A.a.* σ^{54}(16-41), and *A.a.* σ^{54}(WC-16-41) were placed into a pET His6 MBP TEV LIC cloning vector (1M) obtained from the UC Berkeley MacroLab. His-MBP-AID was expressed by growing *E. coli* Rosetta cells at 37°C in 1L of isotope-labeled M9 minimal media to an optical density of 0.6 and induced with 0.5 mM isopropyl thiogalactopyranoside then harvested after 8-12 hours. Leucine, valine labeled peptide was obtained by growth on 1L $^{2}H_2O$ M9 media with deuterated glucose

and the addition of 100 mg of the Leu-Val precursor α-ketoisovalerate added 1 hour before induction. This produces u-[^{2}H,^{12}C], δ1,δ2-[^{1}H,^{13}C] resulting in NMR active nuclei on the δ carbon of the leucine side chains and on the γ carbon of the valine side chains. The expressed protein was found in inclusion bodies after sonicating and pelleting the lysates. Proteins in the pelleted inclusion bodies were unfolded in Denaturing Wash Buffer (8 M urea, 20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4), residual insoluble protein was pelleted and then soluble supernatant was passed through a 0.2 μm filter before loading onto a NiNTA column. Samples were washed and refolded on the column by passing through 5 column volumes of Ni Wash Buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4), then eluted with Ni Elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). His-MBP-AID was dialyzed into 20 mM Tris, 1 mM DTT, 0.1 mM EDTA, pH 7.4 buffer overnight, transferred to a falcon tube with the addition of 20% glycerol. This protein was cut with TEV protease and passed through 30k and 10k MWCO Amicon Ultra centrifugal filters, after which the majority of cleaved AID peptide (3.5-4.5 kDa) remained in the flow through. Peptide was reconcentrated into NMR buffer (20 mM Tris, 200 mM KCl, pH 7.0) in a 3 kDa MWCO Amicon Ultra centrifugal filter.

NtrC1^{C}(121-387) was cloned into pET28a vectors with an N-terminal, TEV protease cleavable His tag. NtrC1^{C}(E239A) was prepared using QuikChange site directed mutagenesis on the NtrC1^{C} plasmid. Protein was expressed in either Luria-Bertani media or $^{2}H_2O$ M9 minimal media, grown to an optical density at 600 nm of 0.6, induced with 0.5 mM isopropyl thiogalactopyranoside. Pelleted cells were sonicated and heated at 75°C for 20 minutes, which precipitated most of the proteins other than the thermophilic *A.a* NtrC1^{C}. Protein was purified on an Ni-NTA column and dialyzed overnight back into Ni Wash Buffer with TEV protease. Cut NtrC1^{C} was passed through an NiNTA column to separate it from TEV and the flow through was collected and dialyzed into NMR Buffer (20 mM Tris, 200 mM KCl, 5% glycerol pH 7.0).

The RNAP(ααββ'ω) expression plasmid was a gift from the Buck Lab (Wigneshweraraj et al. 2003). RNAP protein complex was expressed in cells grown on Luria-Bertani media at 37°C to an optical density at 600 nm of 0.6. Cells were cold shocked on ice for 30 minutes before induction with 0.5 mM isopropyl thiogalactopyranoside and then grown for 6-8 hours at 25°C. The RNAP β subunit contains a C-terminal His tag, and all subunits of the complex were eluted together on an NiNTA column, then dialyzed into low salt buffer and further purified using a heparin column eluted with an NaCl gradient.

2.5.2 Preparation of the ATP analog ADP-BeF_3^-

ADP-BeF_3^- analog was prepared with a 1:1:8:1 ratio of ADP:$BeCl_2$:NaF:$MgCl_2$ by thawing 0.1 M ADP to room temperature and mixing it with NMR buffer to a final concentration of 20 mM. $BeCl_2$ was added to a final concentration 20 mM causing a precipitate to be observed. The precipitate disappears when NaF was added to a final concentration of 160 mM. Finally, slight excess of $MgCl_2$ was added to achieve a final concentration of 25 mM and precipitate was observed again. The solution was passed through a 0.2 μm filter to remove remaining precipitates.

2.5.3 NMR spectroscopy of σ^{54} and RNA Polymerase

Samples were obtained by mixing excess RNAP with ^{15}N-labeled σ^{54} constructs at low concentration and concentrated to 25 μM as a complex with a 10k MWCO Amicon Ultra centrifugal filter. NMR data were collected at 298 K on Bruker Avance 800 MHz or 600 MHz spectrometers. Chemical shift changes were observed with a ^{1}H-^{15}N HSQC. Data were processed with NMRPipe (Delaglio et al. 1995) and chemical shift analysis was undertaken with CARA (Keller 2004) or MestreNOVA (Cobas and Sardina 2003).

2.5.4 NMR spectroscopy of σ^{54}(AID) and NtrC1^{C}

Samples were prepared by mixing ^{15}N-labeled, and later deuterated with ^{1}H-^{13}C labeled Leu δ and Val γ, σ^{54} AID constructs with either unlabeled or ^{2}H NtrC1^{C} and premixed ADP-BeF_3^-, ADP alone, or with the ATP-hydrolysis deficient mutant NtrC1^{C}(E239A) and ATP. NMR data were collected at 298 K on Bruker Avance 800 MHz or 600MHz spectrometers. Chemical shift changes of all σ^{54} AID amides were observed using ^{1}H-^{15}N HSQC experiments and chemical shift changes of Leu and Val side chains using ^{1}H-^{13}C methyl-TROSY experiments (Tugarinov et al. 2003). Amide assignments were carried out using 3D ^{15}N-NOESY-HSQC experiments to both identify the amino acid type of amides by their side chain shifts and connect amides to their neighbor Hα resonances with the sequential assignment approach (Wüthrich 1986). Data were processed with NMRPipe (Delaglio et al. 1995) and assignment analysis used the programs CARA (Keller 2004) or MestReNova (Cobas and Sardina 2003).

2.5.5 Fluorescence Anisotropy, Native Gels, and Size Exclusion Chromatography of σ^{54} and NtrC1^{C}

A tryptophan and a cysteine residue were introduced in front of the start site of the σ^{54}(16-41) construct by QuikChange site directed mutagenesis. Expression and purification techniques were the same as σ^{54}(16-41). Purified peptide was incubated with excess Alexa Fluor 488 maleimide which attached the dye to the single cysteine residue. Reactions were kept in the dark and excess dye was removed by reconcentrating the peptide in a 3k MWCO Amicon Ultra centrifugal column. 0.16 μM concentration of Alexa488-σ^{54}(16-41) (or 5 μM of Alexa488-full length σ^{54} (K2C)) was mixed with varying concentrations of NtrC1^{C} and either ADP or the ATP analog ADP-BeF_3^-. Fluorescence anisotropy was measured on a DTX880 (Beckman Coulter) plate reader. Samples were also run on a native PAGE (4-15%) gel in running buffer that included ADP-BeF_3^- or ADP and visualized with a Typhoon (GE Life Sciences) to look for dye fluorescence. Size-exclusion chromatography was performed by mixing concentrated σ^{54} with excess NtrC1^{C} in the presence of ADP-BeF_3^- or ADP. Samples were injected into a Sephadex-75 column and washed with loading buffer containing either ADP-BeF_3^- or ADP. Since the small, σ^{54} AID peptides have little absorbance at 280nm, column fractions were collected and run on SDS-PAGE gels then stained with Coomassie Brilliant Blue to determine the presence or absence of σ^{54} and NtrC1^{C} in each fraction.

Chapter 3: Effects of insertions on σ^{54} activity *in vivo*

3.1 Summary

Bacterial transcription is initiated after RNA polymerase binds to specific DNA sites in the promoter region for the gene. The specificity for particular sets of genes is provided by the sequence preference of the sigma subunit. While most RNA polymerase holoenzymes are able to open these promoter sites immediately, polymerase holoenzymes formed with members of the σ^{54} family must first be acted on by a transcriptional activator before they can initiate transcription. Such activators also bind sequence specifically upstream of the promoter. A conserved AAA+ ATPase domain in the activators drives their assembly into hexamer rings, which then bind to the N-terminal activator interacting domain of σ^{54}. ATP hydrolysis by the activator drives a conformational change, moving conserved loops near the center of the ATPase ring. The motion of these loops causes a conformational change in σ^{54} that enables the RNA polymerase holoenzyme to open DNA. To provide insight into the mechanism by which the ATPase acts, we measured the *in vivo* activity of σ^{54} variants with insertions or deletions in the region connecting the activator interaction domain (AID) to the core of σ^{54}. The data show that a flexible region of amino acids is necessary between the site of activator interaction and the core of σ^{54} for full activity. These results suggest a mechanism in which the activator threads the AID and portions of the linker domain through the central pore of the transcriptional activator. The force generated by the ATPase would drive conformational changes in the RNA polymerase holoenzyme that enable it to melt the promoter DNA and initiate transcription.

3.2 Introduction

Sigma factors are modular subunits that associate with bacterial core RNA polymerase to form the RNA polymerase holoenzyme. The sigma factors are the primary determinant of promoter specificity for polymerase. While core RNA polymerase ($\alpha_2\beta\beta'\omega$ subunits) is sufficient for transcribing mRNA from melted, single stranded DNA, only the full holoenzyme is able to open DNA at the promoter to initiate transcription (Burgess et al. 1969). The control of sigma factor expression and activity,

as well as access to their conserved binding sites in genomic DNA, are frequently elements of bacterial regulation of gene expression (Losick and Pero 1981).

Two distinct classes of sigma factors exist, with no sequence homology between them, which bind to and act with the same core RNA polymerase enzyme. The σ^{70} class has many members, with representatives always present in bacteria (Paget and Helmann 2003). The σ^{70} variants have distinct promoter binding specificities and considerable variation in size (Lonetto, Gribskov, and Gross 1992). The less common σ^{54} class is frequently but not always present in bacteria, and in any given organism there is only one form of σ^{54}, which binds strongly to conserved promoter sites (Buck et al. 2000). RNA polymerase holoenzymes formed with a member of the σ^{70} class are able to initiate transcription immediately after binding to the σ^{70} promoter sites along the DNA (Browning and Busby 2004)(Ghosh, Bose, and Zhang 2010). In contrast, the RNA polymerase holoenzyme formed with a member of the σ^{54} class is able to bind promoter DNA but is not able to initiate transcription until it is acted upon by a transcriptional activator (Buck et al. 2000) (Figure 1.1). The activators bind to enhancer sequences upstream of the promoter, and hence provide specificity for which subsets of genes with σ^{54} binding sites are transcribed under different conditions (Sasse-Dwight and Gralla 1988)(Keener and Kustu 1988). This extra activation requirement gives the cell finely tuned control of genes regulated by σ^{54} and provides a rapid response by allowing assembly of many of the components before the transcriptional activators induce transcription. These features make σ^{54} regulation effective for controlling gene expression in response to stresses, like heat shock and nitrogen starvation (Kazmierczak, Wiedmann, and Boor 2005).

Transcriptional activators always have ATPase domains, and almost always also have DNA binding domains and regulatory domains (Studholme and Dixon 2003). The regulatory regions prevent assembly of the ATPase (thereby keeping it inactive) until a signal is received (Nohaile et al. 1997), with phosphorylation and small molecule ligand binding being the most common signals though many other mechanisms occur (West and Stock 2001). The cellular response to nitrogen limited conditions is one prime example of gene expression controlled by σ^{54}. A histidine kinase (NtrB) autophosphorylates when levels of nitrogen are low in the cell (Keener and Kustu 1988)(Weiss and Magasanik 1988). The kinase then phosphorylates the regulatory receiver domain of NtrC, a transcriptional activator. When phosphorylated the NtrC receiver domain binds the ATPase domain of a neighbor NtrC monomer, and repeats this with other NtrC molecules to assemble into the hexameric ATPase (De Carlo et al. 2006). The enzymatically active hexamer interacts with and acts on the σ^{54}-RNA

polymerase holoenzyme bound at promoters of genes like glnA, coding for glutamine synthetase (Figure 3.1). Glutamine synthetase assimilates ammonia, converting glutamate to glutamine (Tyler 1978).

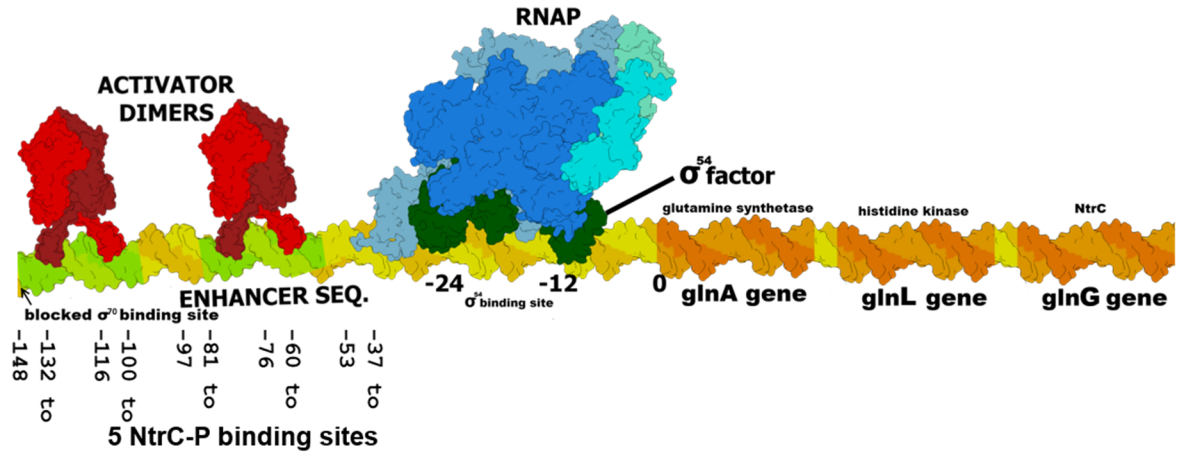


Figure 3.1 Cartoon showing a typical σ^{54}-regulated operon, glnALG. Two NtrC dimers (red) bind to four out of five detected NtrC enhancer sites (lime green) on the DNA in front of the σ^{54} (green) and RNA polymerase holoenzyme (blue) bound at the conserved -24 and -12 sites along the DNA (yellow). After transcription initiation by the activated holoenzyme, RNAP continues to transcribe the glnA gene (orange) as well as genes coding for the histidine kinase that activates NtrC and NtrC itself.

Transcriptional activator ATPase domains are part of a broad class, the ATPases Associated with diverse cellular Activities (AAA+ proteins). Like other AAA+ proteins, the transcriptional activators assemble into hexameric rings with a central pore and hydrolyze ATP to drive a conformational change in their binding partners, in this case σ^{54} (Tucker and Sallai 2007). It is not clear how ATP hydrolysis is coupled to conformational changes in the RNA polymerase-σ^{54} holoenzyme that render it capable of forming the open complex and melting DNA. One proposed mechanism is that the transcriptional activators contact the σ^{54} N-terminal activator interaction domain (AID) and also the RNA polymerase holoenzyme near the site of DNA opening, and through ATP hydrolysis in some way help melt the DNA (Yang et al. 2015). However, this is difficult to reconcile with the tight binding of the transcriptional activators to the σ^{54} AID alone (Siegel et al., submitted). To probe possible features for the transduction mechanism, we examined the activity of mutant forms of σ^{54} with insertions or deletions in the N-terminal region. The data suggest that a 'threading' mechanism, common to many AAA+ ATPases, is more consistent with the experimental observations. In this model, successive rounds of ATP hydrolysis by the activators pull the N-terminal AID of σ^{54} through its central pore and the force generated by this process is applied to the σ^{54}-RNA polymerase holoenzyme, driving a conformational change that enables RNA polymerase to form melt DNA to form the open complex.

3.3 Results

To study the mechanism of σ^{54} activation, we prepared an *in vivo* system starting with an *E. coli* strain from the Keio collection (Baba et al. 2006) with the *rpoN* gene coding for σ^{54} knocked out (ΔrpoN::kan cells). These cells were able to grow in a rich medium, but could not grow in a minimal medium. Under the nitrogen limited conditions of minimal media, *E. coli* requires functional σ^{54} to initiate transcription of many genes involved in nitrogen uptake and fixation (Gyaneshwar et al. 2005), including the glnA gene coding for glutamine synthetase, which converts glutamate and ammonium to glutamine (Tyler 1978). Glutamine synthetase is essential for 'fixing' ammonia, the sole nitrogen source in minimal media, as glutamine, and therefore σ^{54} is essential for growth. We tested the functionality of variant forms of σ^{54} (Figure 3.2) with insertions or deletions of domains in the N-terminal region by introducing plasmids containing the desired σ^{54} sequence and measuring cell growth in minimal medium.

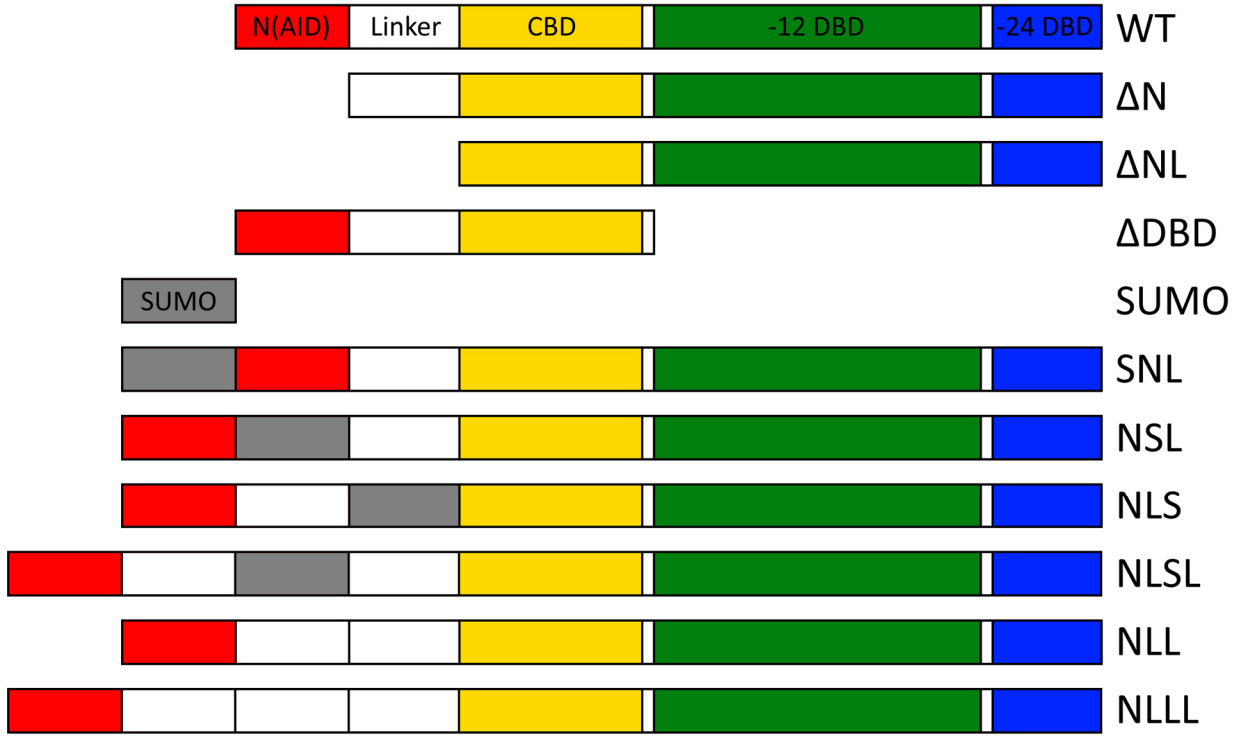


Figure 3.2 Domain architecture of the σ^{54} insertion and deletion mutants used in growth assays. A small, stably folded SUMO domain (gray) was inserted before, after, or in between the AID (red) and linker (white). The CBD (yellow), -12 DBD (green), and -24 DBD (blue) were generally left unchanged. Shorthand notation to refer to each mutant construct is listed on the right based on the ordering of the AID (N), SUMO (S) and Linker (L) domains.

3.3.1 *In vivo* growth of cells with σ^{54} domain deletion mutants

Cells transformed with WT-σ^{54} grew well on M9 medium, while cells lacking any functional σ^{54} showed only extremely slow growth. This verifies the effectiveness of the growth assays in distinguishing between fully functional and fully non-functional σ^{54}. When ΔNL-σ^{54} or ΔDBD- σ^{54} were the only forms of σ^{54} available, the cells grew no better than they did without any σ^{54}. As found in previous work, both the N-terminal activator interacting domain and C-terminal DNA binding domains are absolutely essential to σ^{54} activity (Figure 3.3).

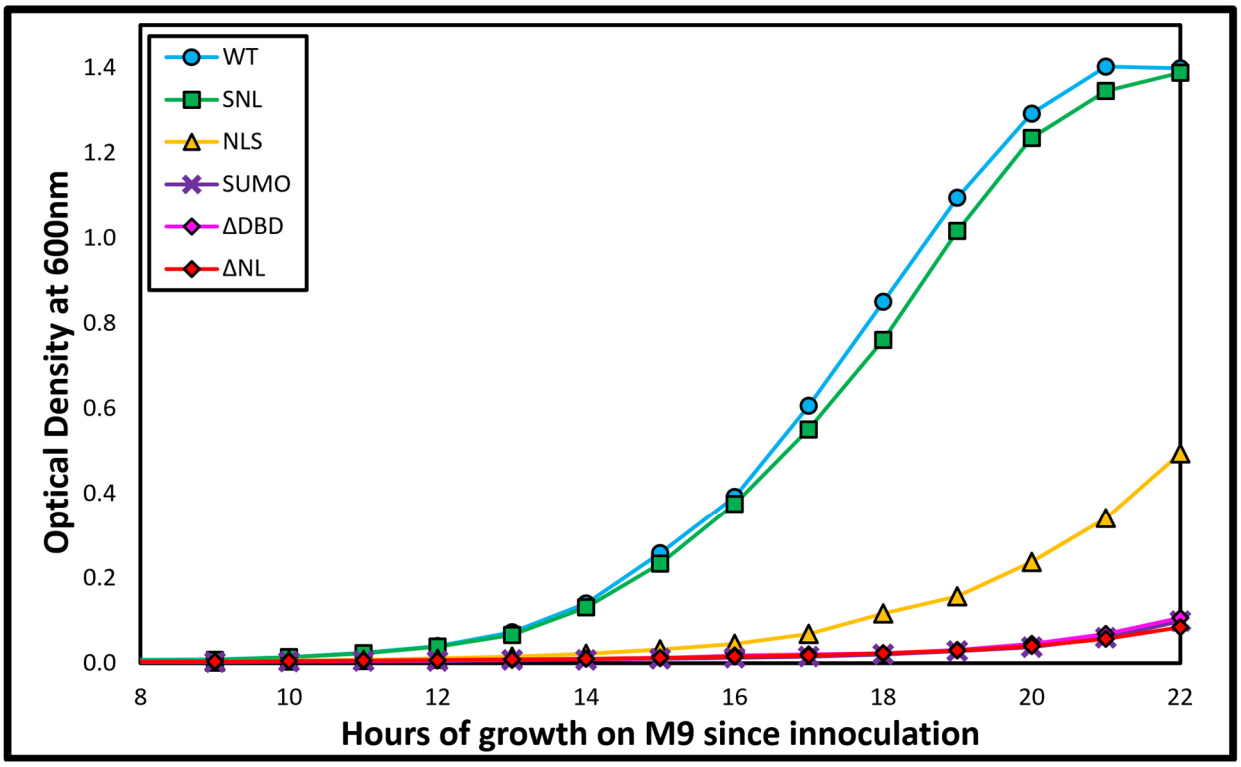


Figure 3.3 Growth on minimal media of *E. coli* ΔrpoN cells from the Keio collection containing plasmids conferring kanamycin resistance and a mutant version of σ^{54}. Introduction of WT σ^{54} or σ^{54} with SUMO fused at the N-terminus (SNL) generated similar growth rates, but introduction of σ^{54} with SUMO inserted between the linker domain and CBD (NLS) grew at a slower rate. Plasmids coding for no σ^{54}, σ^{54} without the AID or linker (ΔNL), or σ^{54} with only the AID, linker and CBD (ΔDBD) did not grow.

3.3.2 *In vivo* growth of cells with σ^{54} containing bulky domains

To investigate steric constraints on the N-terminal region, the well-folded 101 amino acid domain, SUMO (Sheng and Liao 2002), was introduced before and after the activator interacting domain and linker (SNL and NLS respectively) to probe the activation mechanism. As shown in Figure 3.3, while SNL-σ^{54} grew as well as WT-σ^{54},

the NLS-σ^{54} exhibited an intermediate growth rate, indicating it has lower transcriptional activation activity than WT-σ^{54} but higher activity than σ^{54}-ΔNL or absence of any σ^{54}.

The same assays were done in M9 medium using BL21(DE3)pLysS cells with their rpoN gene coding for σ^{54} removed and variant forms of σ^{54} introduced on plasmids. Cells with WT-σ^{54} and SNL-σ^{54} again grew at essentially the same rate, while cells with NLS-σ^{54} again grew at an intermediate rate. However, cells without σ^{54} or with ΔNL-σ^{54} did not grow at all (Figure 3.4). The growth rates showed an even more pronounced difference between cells with functional σ^{54} and those without than the Keio strain. In the BL21 ΔrpoN cells, each introduced plasmid conferred the same phenotype as the Keio cells: normal growth for WT and SNL, intermediate growth for NLS, and no growth for ΔNL and the plasmid only control. When these cells were induced with IPTG to overexpress σ^{54}, none of the introduced plasmids could restore growth, indicating that overexpressed functional σ^{54} has as much of a negative impact on cell growth as non-functional σ^{54}.

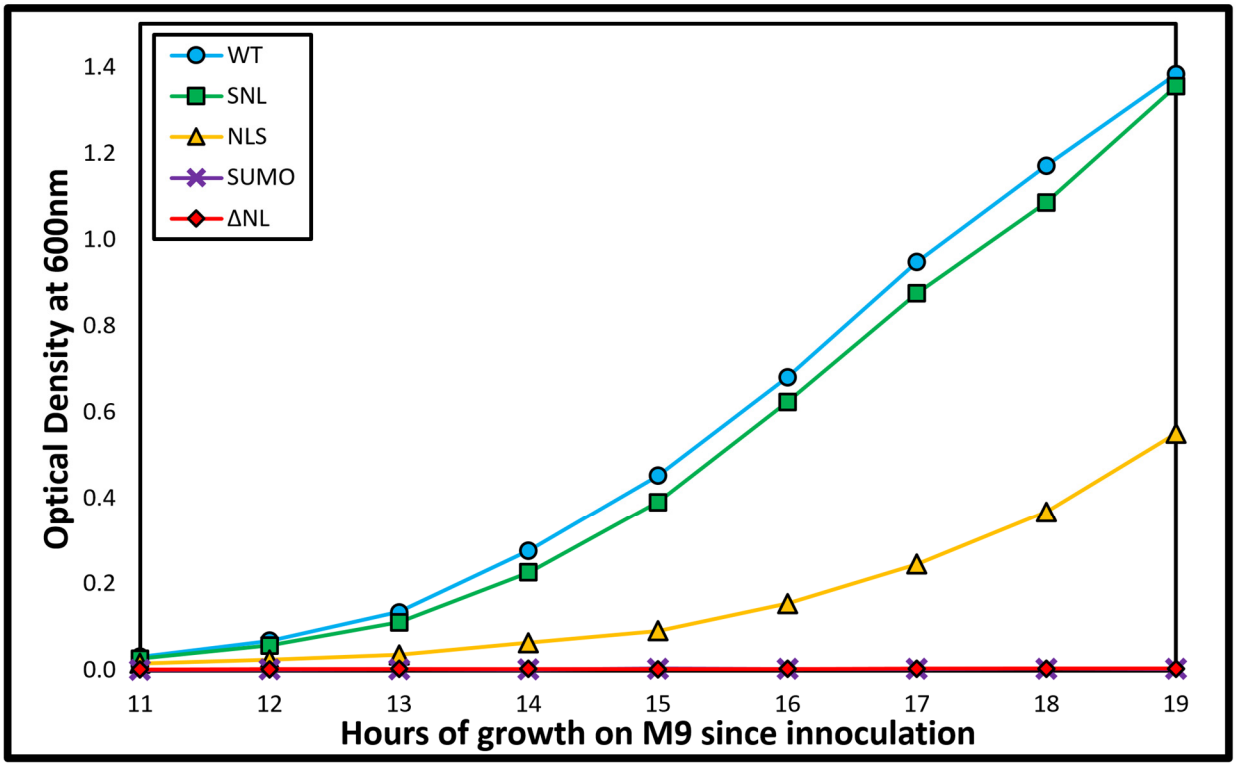


Figure 3.4 Growth on minimal media of *E. coli* BL21(DE2)pLysS ΔrpoN cells containing plasmids conferring kanamycin resistance and a mutant version of σ^{54}. Plasmids coding for WT σ^{54} and σ^{54} with SUMO fused at the N-terminus (SNL) grew at similar rates, but grew slower when SUMO was inserted between the linker domain and CBD (NLS). Plasmids coding for no σ^{54}, or σ^{54} without the AID or linker (ΔNL) has no measurable growth. When cells with any of these plasmids were induced with IPTG they did not grow (not shown).

3.3.3 *In vivo* activity cells with σ^{54} variants

To complement the growth rates, we implemented an activity assay to measure the relative levels of functional σ^{54} over the cell's lifetime to confirm that the slow growth phenotypes represent a loss of σ^{54} function. Plasmids carrying a kanamycin resistance gene and containing both a variant form of σ^{54} and a constitutively active form of the activator, NtrC(D54E, S160F) (Klose, Weiss, and Kustu 1993)(Nohaile et al. 1997) under control of the T7 promoter for use in autoinduction, were transformed into BL21(DE3)pLysS ΔrpoN cells. These cells were also transformed with an additional plasmid carrying the upstream region of the glnHp2 promoter, which includes five NtrC binding sites and the -12 and -24 σ^{54} promoter sequences upstream of the gene start site (Claverie-Martin and Magasanik 1991). The glnH gene on these plasmids was replaced with the gene lacZ coding for β-galactosidase.

Cells were grown in auto-induction media (Studier 2014) which automatically induces expression when the media runs out of glucose late in the log phase, ensuring that induction occurs at roughly the same optical density of cells even if different cell cultures reach that point at different times. 20 hours after the initial inoculation, the cells were pelleted and their contents were assayed for β-galactosidase activity by colorimetric measurements of the yellow ortho-nitrophenyl cleavage product of ortho-Nitrophenyl-β-galactoside (ONPG). As a positive control, we introduced a plasmid containing WT-σ^{54} along with NtrC(D54E, S160F), and as a negative control we introduced a plasmid containing just NtrC(D54E,S160F).

The variants NLL-σ^{54} and NLLL-σ^{54}, which contain two and three repeated linker regions in a row respectively rather than the single linker region in WT-σ^{54}, resulted in activity as high as WT-σ^{54}. However, when one of these linker domains is replaced with a roughly similarly length, but folded, SUMO domain, as in NLSL-σ^{54} and NSL-σ^{54}, we observed a significantly lower activity, falling in between WT-σ^{54} and no σ^{54}. The folded domain did not affect the activity when placed at the very N-terminus in front of the activator interacting domain, as in the case of SNL-σ^{54}. When the folded domain was placed close to the neighboring core binding domain without a linker region for spacing, as in NLS-σ^{54}, activity levels were as low as in the absence of σ^{54}. Similarly, σ^{54} mutants lacking the N-terminal activator interacting domain, ΔN-σ^{54} and ΔNL-σ^{54}, showed no σ^{54} activity (Figure 3.5).

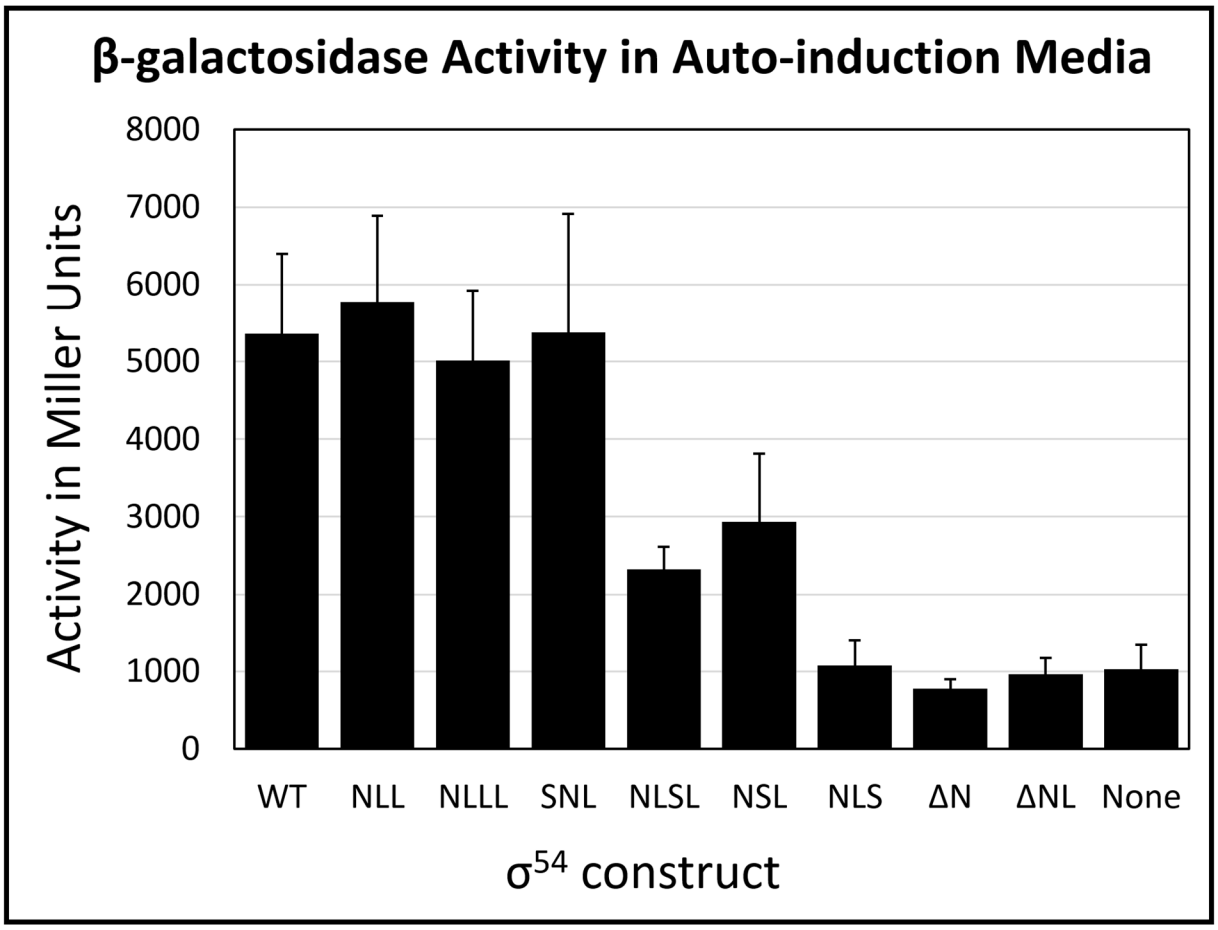


Figure 3.5 Relative β-galactosidase activity due to transcription of lacZ behind the plasmid-born glnHp2-lacZ promoter. BL21(DE3) ΔrpoN cells were transformed with different σ^{54} insertions and deletions and grown in auto-induction media overnight. Equal cell counts were used to determine the amount of σ^{54}-dependent β-galactosidase expression by measuring the rate of cleavage of ONPG into galactose and the yellow product o-nitrophenol as determined by absorbance at 420nm. Trials were performed in triplicate. Three levels of σ^{54} activity can be observed for each of the σ^{54} insertion and deletion mutants: regular (WT, NLL, NLLL, SNL), impaired (NLSL, NSL), and background levels (NLS, ΔN, ΔNL, None).

Analogous experiments were repeated in Luria-Bertani medium, either with induction of expression with 200 μM IPTG during the mid-log growth phase, or left to grow the same amount of time without induction. Reporter expression was higher with IPTG induction than it was with autoinduction but had more variability, perhaps because slight differences in the point during the log growth phase when the cells are induced can result in greater differences in σ^{54} protein yield in each trial. Though the variability is higher, the relative level of reporter expression with different σ^{54} variants still follows the same order as it did in the auto-induction medium, with SUMO insertions after the N-terminal segment leading to substantially reduced but not entirely eliminated reporter expression. Relative to the auto-induction experiments, the wild type σ^{54} was slightly higher in activity than the other constructs. All variants showed greatly reduced levels

of activity when cells were not induced with IPTG, though low levels of activity were seen, which we attribute to leaky expression (Figure 3.6).

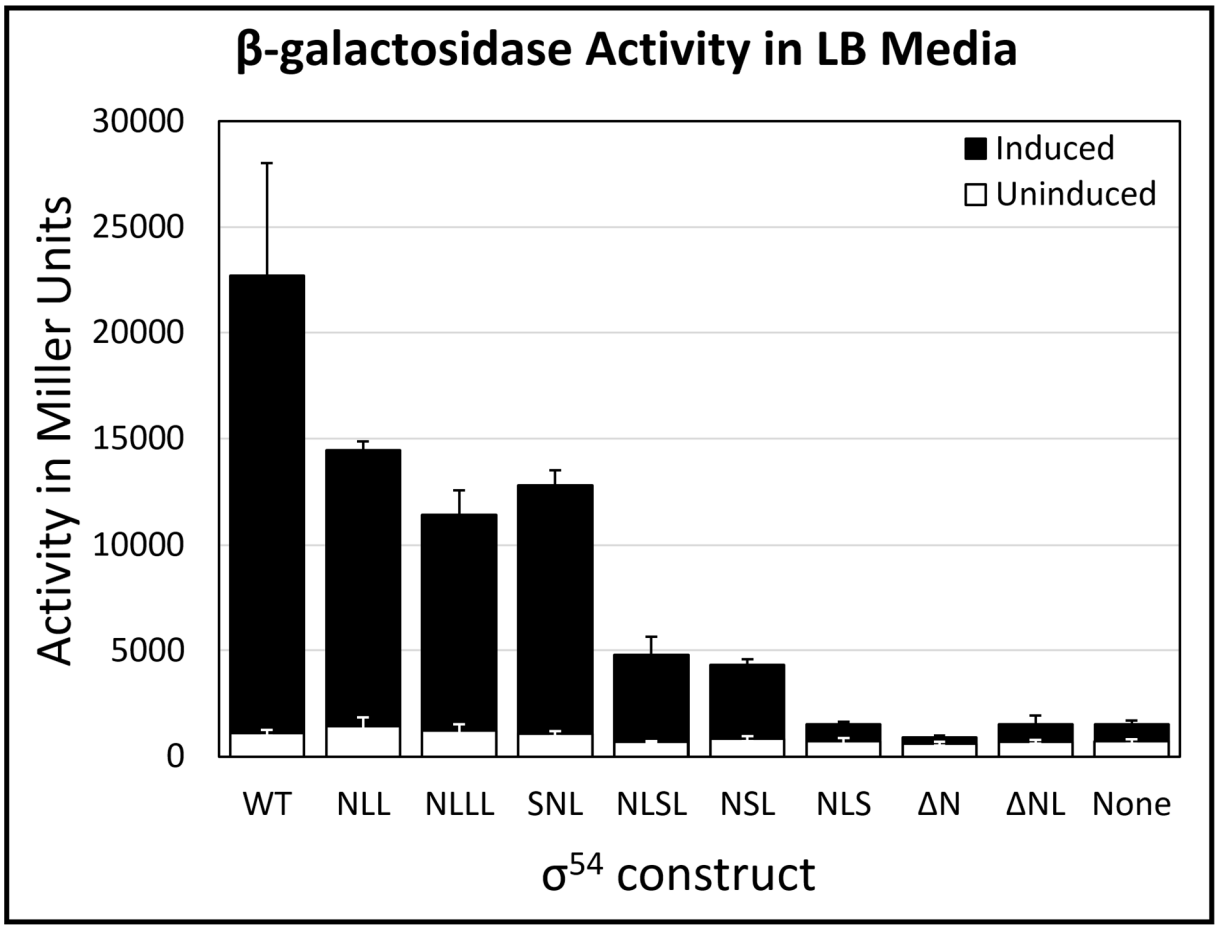


Figure 3.6 Relative β-galactosidase activity due to transcription of lacZ behind the plasmid-born glnHp2-lacZ promoter. BL21(DE3) ΔrpoN cells were transformed with different σ^{54} insertions and deletions and grown Luria-Bertani media induced with IPTG during the mid-log growth phase (black) or left uninduced for the same amount of time (white). Equal cell counts were used to determine the amount of σ^{54}-dependent β-galactosidase expression by measuring the rate of cleavage of ONPG into galactose and the yellow product o-nitrophenol as determined by absorbance at 420nm. Trials were performed in triplicate. Three levels of σ^{54} activity can be observed for each of the σ^{54} insertion and deletion mutants: regular (WT, NLL, NLLL, SNL), impaired (NLSL, NSL), and background levels (NLS, ΔN, ΔNL, None). Uninduced cultures all had similarly low background levels of activity.

3.4 Discussion

Bacterial σ^{54}-polymerase requires the action of a transcriptional activator to initiate transcription, with regulation of activator assembly serving as the primary control. Activators contact the N-terminal AID of σ^{54} and must hydrolyze ATP to enable transcription initiation (Chaney et al. 2001), but a mechanistic understanding of these

events and how they reconfigure the polymerase-DNA complex into a form capable of opening DNA is not understood. To probe this mechanism, we measured the effectiveness of *E. coli* with modified versions of σ^{54} to grow in nitrogen limited media or to initiate transcription with the constitutively active activator, NtrC(D54E, S160F).

3.4.1 Proposed Mechanisms of Transcription Activation

The first step in the process of activating σ^{54}-polymerase for transcription is the formation of an encounter complex between conserved GAFTGA loops from the assembled activator ATPase domains and the N-terminal activator interacting domain of σ^{54}. The encounter complex forms with substantial affinity only when the activators are primarily in their ATP state (Chaney et al. 2001). Cryo EM densities of the ATPase rings in ADP and ATP bound states using a variety of non-hydrolyzable ATP analogs showed that the GAFTGA loops are raised further above the ring with ATP bound than with ADP (Chen et al. 2007), with the transition viewed as a possible mechanism for driving the conformational change in the polymerase. High resolution crystal structures of the $NtrC1^{C}$ uniformly bound to ADP (Lee et al. 2003) and the ATP hydrolysis deficient mutant $NtrC1^{C}$(E239A) uniformly bound to ATP (Chen et al. 2010) demonstrate a similar raising and lowering of the GAFTGA loops dependent on the nucleotide state (Figure 3.7). A recent structure was determined in the presence of ADP-BeF_3^- to mimic ATP, which revealed variable states of the GAFTGA loops ranging between the raised and lowered conformations (Sysoeva et al. 2013). Consistent with previous work, subunits with the most ATP-like sites (with BeF_3^- correctly positioned to mimic the third phosphate of ATP) have the loops raised, while the more ADP-like states (with BeF_3^- less tightly bound or not present) have the loops lowered. The spiral nature of the loops is reminiscent of the Rho helicase structure (also a AAA+ ATPase), which is believed to drive the single stranded nucleic acid through a central pore (Skordalakes and Berger 2006).

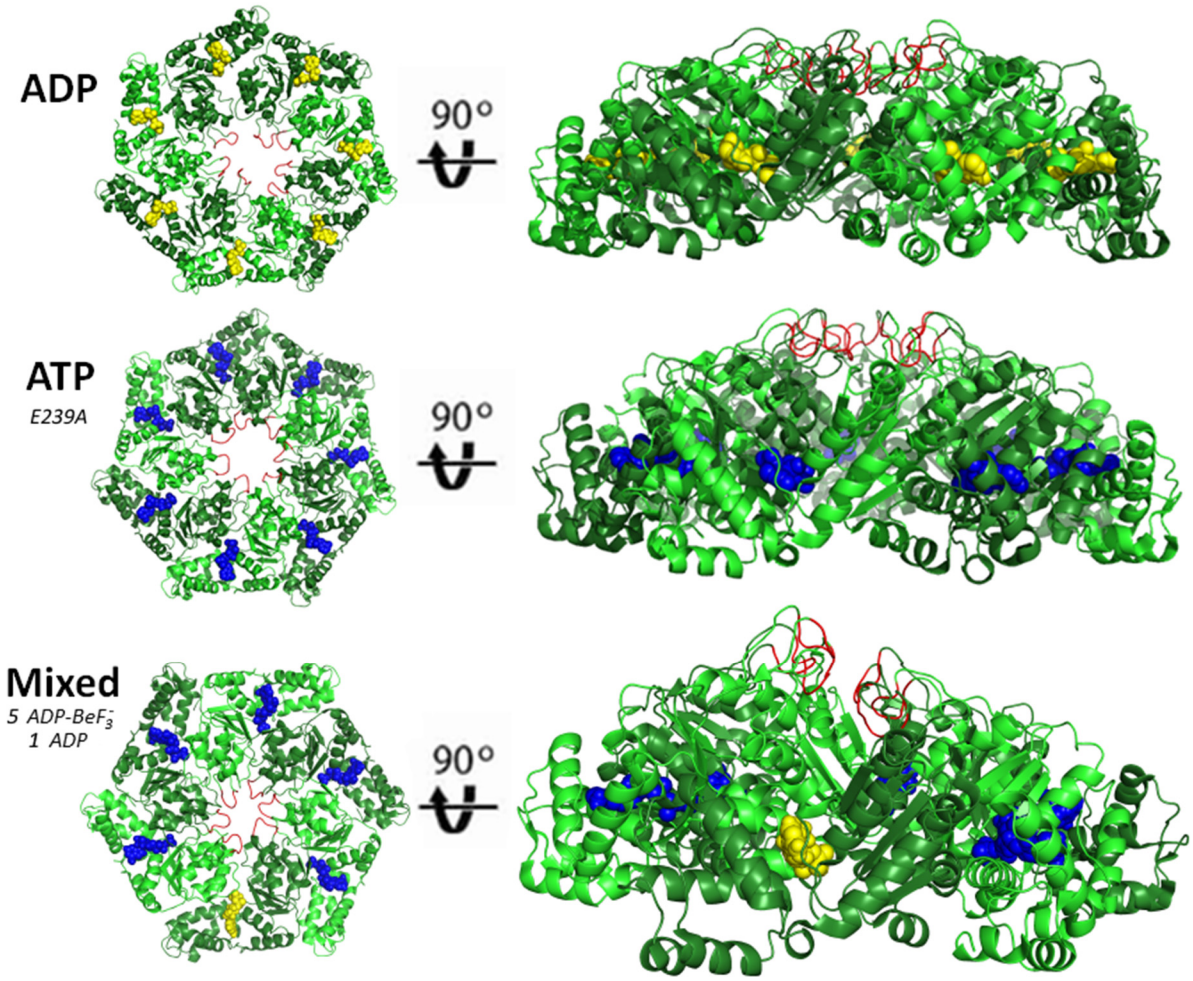


Figure 3.7 Crystal structures showing the NtrC1^{C} heptamer (green) uniformly occupied with ADP (yellow), the ATP hydrolysis deficient mutant NtrC1^{C}(E239A) heptamer uniformly occupied with ATP (blue), and the NtrC1^{C} hexamer bound to a mixture of nucleotides ADP (yellow) and ADP-BeF_3^- (blue). The location of the conserved GAFTGA loops are highlighted (red). A side view of each structure shows the raising of the GAFTGA loops in the ATP-like nucleotide state. PDB from top to bottom: 1NY6, 3M0E, 4LY6.

In one proposed mechanism for the action of the ATPase on σ^{54}-polymerase, the AID forms an interface with other parts of σ^{54} and perhaps other polymerase subunits and/or DNA to bind the activator ATPase (Yang et al. 2015). The motion of the GAFTGA loops within this interface, act on polymerase during ATP hydrolysis, inducing melting of the DNA at the -12 site. We will refer to this as a bind and release mechanism (Figure 3.8).

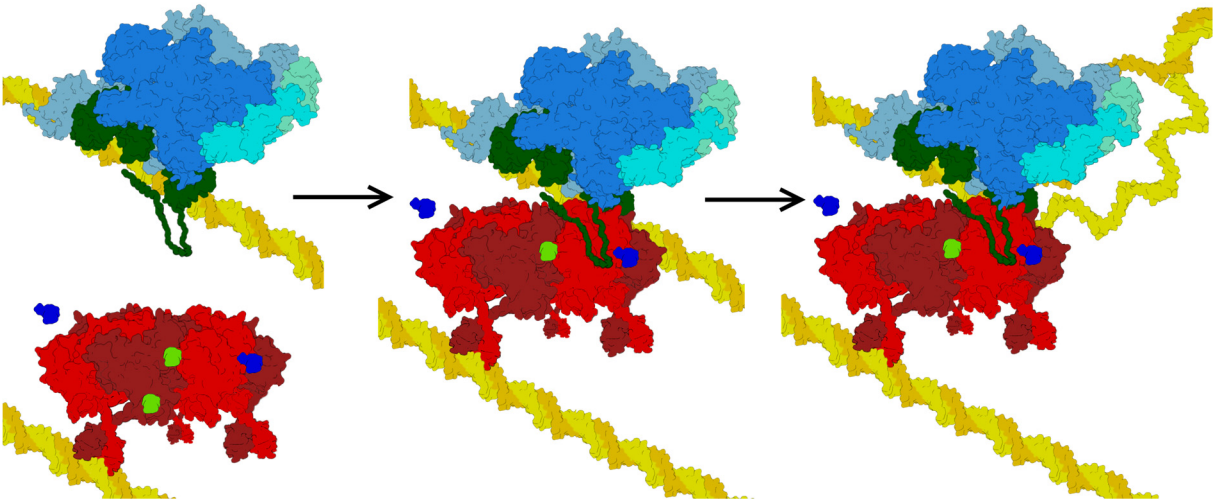


Figure 3.8 Diagram of σ^{54} transcription initiation using a hypothesized bind and release mechanism. (1) σ^{54} (green) is bound to core RNAP (blue) and assembles on the DNA (yellow). The helical σ^{54} AID blocks initiation of transcription by binding to the -12 DBD and inhibiting open complex formation. (2) When a transcriptional activator (red) is activated it binds the σ^{54} AID and pulls it away from other σ^{54} domains. (3) Without the inhibitory effect of the AID, σ^{54} is able to form the open complex with DNA.

Our findings that the AID is flexible and that some linker insertions do not adversely affect function while others do, suggest an alternate model. The initial binding of the σ^{54} AID to the GAFTGA loops, combined with the movement of the loops as the ATPase hexamer goes through ATP hydrolysis, are consistent with a mechanism in which the AID is pulled through the activator to drive conformational changes that enable opening of promoter DNA. We will refer to this as a 'threading' mechanism (Figure 3.9).

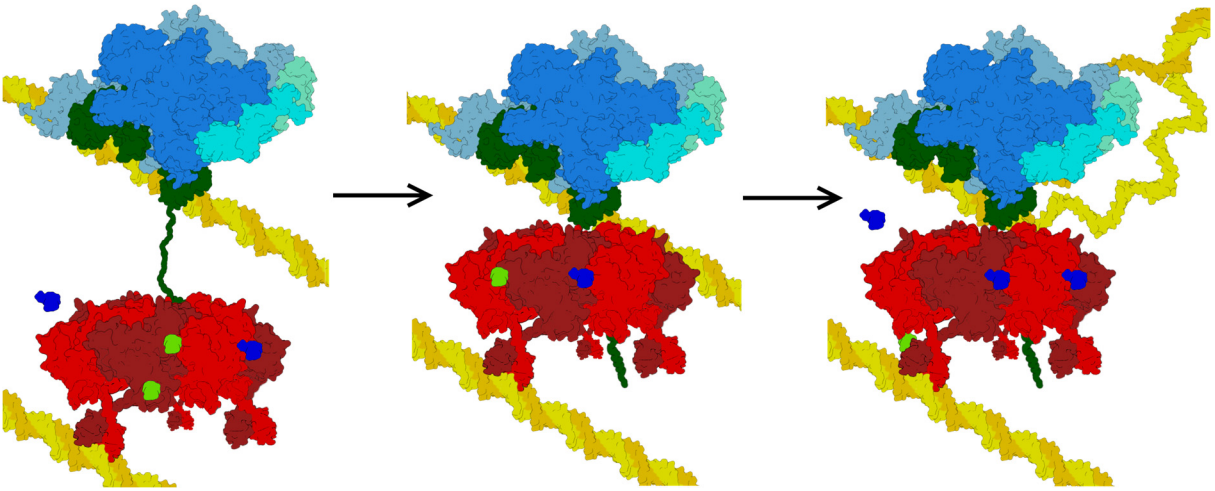


Figure 3.9 Diagram of σ^{54} transcription initiation using our proposed threading mechanism. (1) σ^{54} (green) is bound to core RNAP (blue) and assembles on the DNA (yellow) but the σ^{54} AID remains free and unstructured. (2) When a transcriptional activator (red) is activated it binds the σ^{54} AID and threads it through its central pore through multiple rounds of ATP hydrolysis. (3) This threading force drives a conformational change in the rest of σ^{54} and the RNAP holoenzyme that causes the formation of the open complex with DNA.

3.4.2 The role of the N-terminal activator interacting domain

σ^{54} proteins lacking the first 47 residues from the N-terminal AID cannot initiate transcription. The lack of function that we observe for the N-terminal deletion is consistent with previous truncation studies, which also found that the central part of the AID is necessary for initiation, while the first ≈ 16 residues may impact activation but do not affect binding to the site of DNA opening (Wong, Gralla, and Tintut 1994). The fact that deletion of the AID prevents activation indicates that the AID does not function as an inhibitor of σ^{54} which prevents the σ^{54}-RNAP holoenzyme from opening the promoter until it is modified by the transcriptional activators.

Between the N-terminal AID and the remaining domains of σ^{54} is a linker region with very low sequence conservation, and no predicted secondary structure (Southern and Merrick 2000). This segment varies in length across species from around ≈ 20 amino acids in *A. aeolicus* to ≈ 60 in *E. coli* σ^{54}. When we increased the length of the linker by adding one or two additional copies in series, the *in vivo* activity assays showed that these insertions have little influence on the functionality of σ^{54}. This is consistent with a model in which the N-terminal region is only weakly associated to the RNAP holoenzyme, remaining flexible enough to contact and hold the activator ATPase ring in the vicinity of polymerase (Siegel et al., submitted).

The structure of the ATPase domain from the *A. aeolicus* transcriptional activator NtrC1 determined in the presence of the ATP analog ADP-BeF_3^- (Sysoeva et al. 2013), has the GAFTGA loops forming a spiral array around a central pore. Although the GAFTGA loop has been implicated as a direct contact to σ^{54}, there is no detailed information available about the interaction. However, given the strong similarity to Rho helicase (Skordalakes and Berger 2006) it seems plausible that motion of the loops upon ATP hydrolysis could pull the AID into the pore of the ATPase hexamer. Our NMR studies show that residues 16-41 are immobilized upon interaction. Though we do not know the orientation of the bound σ^{54} segment, the central position of the contacted region in the AID suggests that it might be pulled into the pore as a loop.

3.4.3 Insertion of a folded domain into the linker inhibits σ^{54} activity, but an N-terminal addition does not

If the transcriptional activator threads the N-terminal region of σ^{54}, then we reasoned that insertion of a folded domain could interfere with its function. To test this, we chose SUMO as a small, stably folded unit to insert into σ^{54}. The addition of SUMO to the

N-terminus of σ^{54} had no effect on activity. This observation is compatible with the idea that threading could occur as a loop rather than by the N-terminus entering the pore. The 15 residues between the region of the AID that contacts the ATPase and the N-terminus where SUMO was added could extend to >30Å, probably sufficiently long for the bulky SUMO domain to be out of the way and not interfere with either AID binding to the polymerase (as in the structure) or AID binding to the ATPase domain.

As noted above, insertion of extra copies of the linker connecting the AID and core binding domains did not significantly affect σ^{54} function. However, when the folded SUMO domain was inserted either before the linker or in between two linkers on either end, then function was substantially diminished in both the growth and reporter assays. This is unlikely to be caused by disruption of σ^{54} binding to either the core RNA polymerase or the transcriptional activator because the reduced activity persists even with a full linker domain included on either side of the small folded domain. The linker is sufficiently long that it should allow unperturbed binding to the activator on one side and RNA polymerase on the other side of the folded domain. The reduced *in vivo* activity cannot be explained by a disruption in the spacing of domains of σ^{54} caused by increasing the distance between the AID and CBD, because mutants with double and triple length linkers do not reduce *in vivo* activity despite inserting a similarly large or larger physical distance between the AID and CBD. It is worth noting that the disruption of the σ^{54} linker with a folded domain substantially reduces *in vivo* activity, but does not completely remove it. NLSL and NSL σ^{54} mutants are able to initiate transcription but at a lower level than WT σ^{54}, but still more than the background levels of lacZ transcription observed in cells lacking any σ^{54} entirely.

3.4.4 Consistency with a Threading Mechanism for σ^{54}-Pol activation

In the threading model, movement of the NtrC loops pulls on the σ^{54} N-terminal region during successive rounds of ATP hydrolysis. The ATPase is anchored to the enhancer region DNA, while the σ^{54}-Pol is bound to the promoter region. Often (but not always) there is a binding site for a DNA binding protein, such as IHF, between the promoter and enhancer sites, which brings the activator and the polymerase holoenzyme into closer proximity, and should facilitate the initial binding of the AID to the assembled ATPase (Santero et al. 1992). The N-terminus of σ^{54} is flexible, and the initial ATP hydrolysis would pull it through the activator ATPase ring, but might generate little force (Figure 3.10). As the activator transfers more of the σ^{54} N-terminus through the

ATPase, an increasing force should be applied to the end of the folded core binding domain. The force applied by the ATPase should be maximal if the ATPase contacts the polymerase. Insertion of a folded domain presents an obstacle to the contact between the ATPase and the polymerase, preventing direct contact. This would limit the force generated, and lower the efficiency of driving the conformational change needed to allow the polymerase to open the DNA and hence lower the efficiency of initiation. It is also possible that NtrC may be able to sometimes bypass a folded domain while threading the linker due to the open lock washer structure of the activator's pore (Sysoeva et al. 2013). We note that after the conformational change enabling σ^{54}-Pol to open the promoter has occurred, it is likely that σ^{54} would be released from the activator. Thus unlike processive unfoldases (such as ClpX) the activator should maintain only a loose grip on σ^{54} so that it can be released without completely unfolding it. It seems likely that the energy required to drive the conformational change in the σ^{54}-Pol complex is much less than that required for unfolding a stable protein, so a strong, processive action of the ATPase is not needed.

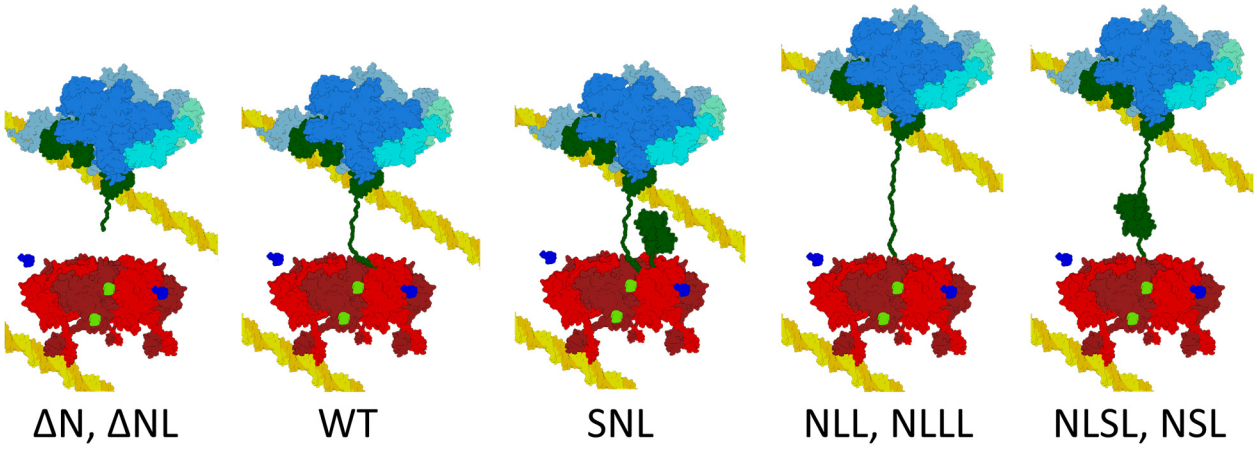


Figure 3.10 Diagram showing the effect that various mutant σ^{54} variants would have on the threading mechanism: (1) σ^{54} without the AID (ΔN, ΔNL) makes no contact with the activators. (2) WT σ^{54} follows the normal threading mechanism. (3) A well-folded domain at the N-terminus (SNL) does not interfere with threading, especially if the AID is threaded as a loop. (4) An extra long but flexible linker (NLL, NLLL) does not disrupt the threading. (5) A well-folded domain in the linker (NLSL, NSL) cannot be easily threaded through the central pore of the transcriptional activator.

3.4.5 Future Directions

The activity studies reported here support a mechanism for σ^{54} transcriptional activation in which the force generated by ATP hydrolysis of the activator drives RNA polymerase to enter the open complex with DNA. However, further experiments are needed to fully establish this mechanism, and to probe the enabling conformational

change in the polymerase. While NMR and X-ray crystallography are valuable for characterizing the structures of domains, these approaches are less useful for probing the more transient structures of intermediate states in the process. As in other systems it should be possible to apply single molecule experiments in this system. Optical tweezer experiments should allow application of force to single σ^{54} molecules in complex with RNA polymerase to probe whether force alone is sufficient for activation, and if so what force is required to drive the conformational change and initiate transcription.

3.5 Materials and Methods

3.5.1 ΔrpoN strains and knockouts

Keio ΔrpoN cells were purchased from the Coli Genetic Stock Center's Keio knockout strain collection (CSGC#: 10404, designation: JW3169-1) F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔrpoN730::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514 (Baba et al. 2006).

ΔrpoN was knocked out of BL21(DE3)pLysS cells using bacteriophage λ with fragments of the Keio cell genome containing the ΔrpoN730::kan insertion (Datsenko and Wanner 2000). BL21(DE3)ΔrpoN cells were selected for kanamycin resistance, indicating a replacement of rpoN with the kanamycin resistance gene. Cells were transformed with pCP20 and grown at 30°C to remove the kanamycin resistance gene (Cherepanov and Wackernagel 1995). Heat-shock competent cells were made by resuspending cells in a solution of 100 mM $CaCl_2$ and 15% glycerol. These cells, which already contain the pLysS plasmid containing T7 lysozyme (p15a, Cam^R) were transformed with 1 or 2 plasmids with compatible origins of replication by heat shock at 42°C for 45 sec followed by incubation and shaking in S.O.C. media for one hour at 37°C. Transformed cells were plated onto LB plates containing the appropriate antibiotics: 50 mg/L kanamycin, 50 mg/L ampicillin, 34 mg/L chloramphenicol.

3.5.2 Plasmid cloning

The σ^{54} insertion and deletion mutants used in the growth assays were cloned into pET-28a vectors (pBR322, Kan^R). For the *in vivo* activity assays, the σ^{54} insertion and deletion mutants were cloned into pCOLADuet-1 vectors (ColA, Kan^R) at the first site and the constitutively active NtrC(D54E,S160F) was cloned into the same vector at the second site. The upstream promoter region of glnH including binding sites for σ^{54} and

NtrC (glnHp2) was cloned into pETDuet-1 (ColE1, Amp^R). The glnH gene itself was replaced with the lacZ gene coding for β-galactosidase. An alternative glnHp1 promoter region (Nohno and Saito 1987) was also removed to ensure that lacZ gene transcription could only be initiated by plasmid-born σ^{54}, which, if functional, will result in high levels of β-galactosidase after a period of cell growth.

3.5.3 Growth assays

Keio ΔrpoN cells and BL21(DE3)pLysS ΔrpoN cells were transformed with the σ^{54} plasmid and grown on M9 minimal media (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 g/L NH_4Cl, 4 g/L glucose) (Maniatis, Fritsch, and Sambrook 1982) for 24 hours. Optical density of the media at 600 nm was recorded beginning 10 hours after inoculation (initial OD_{600nm}=0.0005) at 1 hour intervals to measure cell growth.

3.5.4 Activity assays

BL21(DE3)pLysS ΔrpoN cells were transformed with both the σ^{54}, NtrC(D54E,S160F), Kan^R plasmid and the glnHp2-lacZ, Amp^R plasmid. Cells were then grown in Luria-Bertani medium to the mid-log growth phase (OD_{600nm} = 0.6) and induced with 0.2 mM IPTG for 4 hours at 25°C or grown in auto-induction media (MDASM-5052 Autoclave: 1% tryptone, 0.5% yeast extract, Salts: 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, Sugars: 0.5% glycerol, 0.05% glucose, 0.2% α-lactose) (Studier 2014) for 20 hours at 25°C. Cells were pelleted and resuspended in Z-Buffer (60 mM NaH_2PO_4, 60 mM Na_2HPO_4, 10 mM KCl, 1 mM $MgSO_4$, 1 mM βME, pH 7.0) and diluted to an OD at 600nm of 0.025 and cell membranes were permeabilized by the addition of toluene. Ortho-Nitrophenyl-β-galactoside (ONPG) was added to initiate the β-galactosidase assay and quenched by the addition of 1M Na_2CO_3. β-galactosidase activity was measured by color change of the ONPG at 420nm.

Chapter 4: Force-Dependent σ^{54} Activation

4.1 Summary

In this chapter, I describe work characterizing the effect of force on the *Aquifex aeolicus* σ^{54} core binding domain (CBD). When the structure of the CBD was solved in our lab by NMR, it revealed two distinct subdomains with a small hydrophobic interface between them. We hypothesized that the two CBD subdomains might serve as a fracture point in the σ^{54} structure that drives conformational changes when one of subdomains is disrupted by the application of force. We wanted to determine if the CBD could convert the activation event felt by the N-terminal activator interacting domain into a conformational change in the distant C-terminal DNA binding domains that drives DNA opening. To probe for such a mechanism, we applied force to the ends of the CBD using optical tweezers to look for independent unfolding of one of the two subdomains. While we did notice a short-lived unfolding intermediate, the extension that occurred suggests that it is not a complete unfolding of one of the two subdomains and is most likely an unfolding of the CBD's seventh helix. We found no evidence of a force-dependent conformational change in the CBD on its own, though future experiments applying force to the CBD while it is in complex with the full RNAP holoenzyme may produce more conclusive results.

4.2 Introduction

4.2.1 Structure of the core binding domain

In σ^{70}, the σ_2 domain (*T. thermophilus* 74–254) has been characterized as forming important surface contacts with core RNA polymerase (Vassylyev et al. 2002). Though σ^{54} has no sequence homology with σ^{70}, biochemical analysis has identified a core binding domain (CBD) similarly placed in the sequence (*E. coli* residues 106-269). The σ^{54} CBD shares a number of similarities with the σ^{70} σ_2 domain. Both the σ^{54} CBD and σ^{70} σ_2 domain are primarily helical, with the CBD folding into seven helices (Hong, Doucleff, and Wemmer 2009) and the σ_2 domain folding into eight helices. Both domains have a net negative charge that is important for their interactions with core

RNAP and both contact core RNA polymerase at similar locations as seen in the crystal structure of the σ^{70}-RNAP holoenzyme (Murakami et al. 2002) and in cryo-EM density (Bose et al. 2008) and a recently solved crystal structure (Yang et al. 2015) of the σ^{54}-RNAP holoenzyme.

The NMR structure of the *Aquifex aeolicus* σ^{54} core binding domain (69-198) solved in our lab is composed of two distinct subdomains, a four-helix bundle (60-135) and a three-helix bundle (135-198), with a hydrophobic interface between the two (Hong, Doucleff, and Wemmer 2009) (Figure 4.1). In the absence of the three-helix bundle, the four-helix bundle remains well folded and its structure remains largely unchanged. However, the three-helix bundle requires the surface of the four-helix bundle to fold against and is also capable of folding when the four-helix bundle is present as a separate molecule. The presence of two subdomains in the CBD, held together by hydrophobic interactions and a hinge, suggests the possibility that the domain serves as a conformational fracture point that, subjected to the right conditions, could drive conformational changes to the rest of the σ^{54}-RNAP holoenzyme. Here we discuss the possibility that the CBD is involved in converting the interaction between the transcriptional activators and the activator interacting domain before the CBD's N-terminus into a conformational change in the -12 DNA binding domain after the CBD's C-terminus, in order to drive the formation of the open complex.

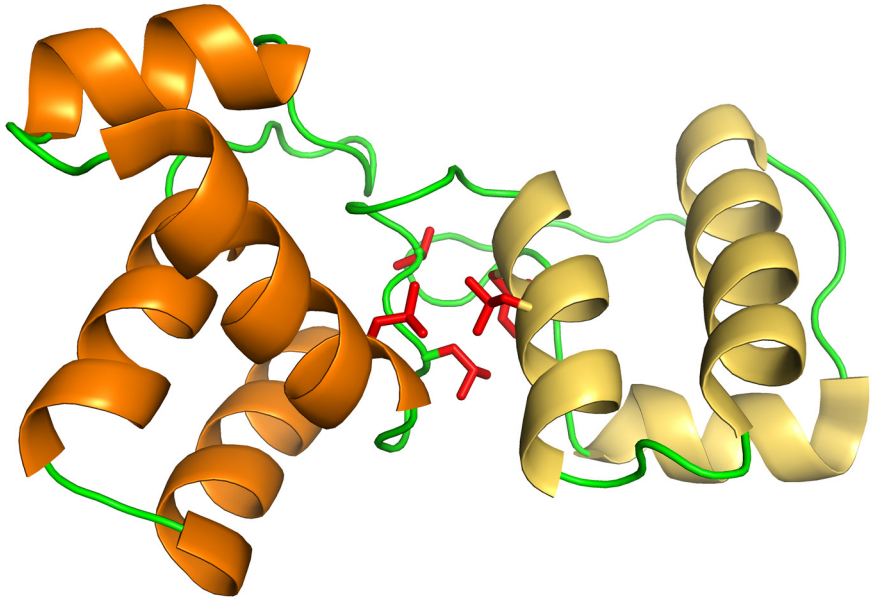


Figure 4.1 NMR Structure of the *A.a.* σ^{54} core binding domain, PDB: 2K9M (Hong, Doucleff, and Wemmer 2009). The four-helix bundle (orange) remains well-folded without the three-helix bundle (yellow). The three-helix bundle only folds in the presence of the four-helix bundle. Side chains of amino acids in the hydrophobic interface (L131, L137, V138, L150) between the two subdomains are shown in red.

4.2.2 Overview of Mechanical Force in Biochemistry

Mechanical forces generated by proteins are important to the mechanisms of many diverse cellular processes including replication, transcription, translation, proteolysis, catalyzed protein folding and unfolding, cargo transport and many others (Bustamante et al. 2004). A common theme among molecular motors is the conversion of energy provided by ATP hydrolysis to drive directed motion of the protein engaged in these processes and apply the force necessary to perform mechanical work. Proteins involved in vesicular transport, like kinesin and dynein, convert the energy of ATP hydrolysis into steps along a microtubule track and provide the force needed to carry their cargo (Howard, Hudspeth, and Vale 1989)(Vallee et al. 2004). DNA and RNA polymerases involved in transcription and translation use the energy from their polymerization reaction to move along their template DNA or RNA (Wang et al. 1998)(Wuite et al. 2000). Helicases apply a mechanical force that unwinds the DNA strands (Delagoutte and von Hippel 2003). Unfoldases apply force to unfold proteins, and translocases pull unfolded polypeptides across membranes (Matouschek 2003). Though structural biology can provide static pictures of these processes, recent advances in single molecule techniques have led to a better understanding of the dynamic mechanisms of these motor proteins.

In this dissertation, we propose a mechanism of σ^{54} transcription initiation in which transcriptional activators hydrolyze ATP to apply a mechanical force to the N-terminus of σ^{54} that drives the formation of the open complex. For this mechanism to make sense, the RNA polymerase holoenzyme must undergo a conformational change brought on by the force of the activator. The interface between the two subdomains of the core binding domain presents one plausible location in the structure where force could bring about a structural rearrangement. We employed optical tweezers to examine the conformational changes and unfolding of the CBD when mechanical forces are applied to either end.

4.2.3 Optical tweezers method

Optical tweezers are used to apply mechanical force to single protein molecules to map the energy landscape of protein folding and study folding intermediates (Smith, S., Cui, Y., Bustamante 2003). One of the simplest and most effective methods of preparing a protein for these experiments involves substituting out native cysteines and replacing the N and C-terminal residues with non-native cysteines (Cecconi et al. 2008). The N- and C-terminal cysteines form two disulfide bonds with thiol-labeled ends of two DNA handles. The other end of each DNA handle is attached to a small molecule with an

affinity towards a specific antibody or protein coated on large polystyrene beads 2 μm in diameter. The beads are held with a pipette on one end and captured with a laser beam on the other (Figure 4.2). In this manner, the protein is held between the two beads but separated from the beads by the DNA handles to avoid any non-specific interactions with them. This allows a mechanical force to be applied to the protein by moving the laser beam and one of the beads along with it, thereby increasing the distance between the two beads attached to either end of the protein. This technique allows the measurement of the force vs. the extension of the protein molecule and has been used to measure intermediates along the folding and unfolding pathway (Cecconi et al. 2005).

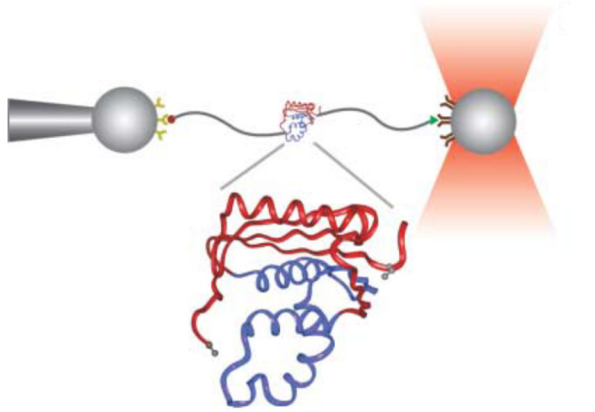


Figure 4.2 Not to scale diagram of an optical tweezer setup showing a folded protein tethered with disulfide bonds on either end to two ≈ 500 bp DNA handles. These handles are bound to two beads coated with proteins or antibodies with an affinity for the other end of the handle. One bead is captured in a laser beam (the 'trap') and the other is held by a pipette. By moving the position of the laser beam and beads, specific forces can be applied to the protein molecule. Adapted from (Cecconi et al. 2005).

4.3 Results

4.3.1 Cysteineless CBD does not affect its structure

The *A.a.* σ^{54} core binding domain (69-198) cysteines at 119 and 140 were mutated to alanines and the residues at 69 and 198 were replaced with cysteines (Figure 4.3). This modified CBD was expressed and purified with an N-terminal His tag and studied to determine whether the removal of native cysteines and introduction of new N-terminal and C-terminal cysteines affected the structure or stability.

```
                                    70          80          90          100
MRGSHHHHHHGSLVPRGSENLYFQGCTPSELEELQQNIKLELEGKEQELALELLNYLNEK

        110         120         130         140         150         160
GFLSKSVEEISDVLRASVEELEKVRQKVLRLEPLGVASKDVWEFLELQIEEIYPEEEEIL

        170         180         190
KKALRDLKRGKKLKPEIKGKLSRLRLFPLSSSAEC
```

Figure 4.3 Sequence of the σ^{54} CBD used in the NMR, denaturation melt, and optical tweezers experiments. The N-terminal His tag (red) was present during the experiments. The two cysteines (blue) at 69 and 198, added at the N-terminus and C-terminus in the structure, were used for the attachment of the handles. The two cysteine to alanine mutations (blue) at 119 and 140 were made to avoid alternate handle attachment sites. The seven helices in the structure are highlighted to show the four-helix bundle (orange) and three-helix bundle (yellow).

Purified WT and cysteineless CBD were subjected to guanidinium chloride dependent unfolding and monitored by circular dichroism (Figure 4.4). The cysteineless mutant of σ^{54} CBD has an m-value of 3.6±1.2 and $\Delta G_{unfolding}$ of 13.4±3.1 kcal/mole, while the WT σ^{54} CBD has an m-value of 3.7±1.4 and $\Delta G_{unfolding}$ of 14.8±5.6 kcal/mole. For WT CBD, this is comparable to the free energy calculated when unfolding is monitored by fluorescence (m-value of 3.5±0.2 and $\Delta G_{unfolding}$ of 14.5±0.7 kcal/mole, not shown). An increase in the number of baseline data points at high guanidinium concentrations for WT CBD and low guanidinium concentrations for cysteineless CBD would be needed to improve the fit. Though there is a slight decrease in stability, replacing the internal cysteines with alanines does not lead to a drastic change in the CBD's free energy of folding.

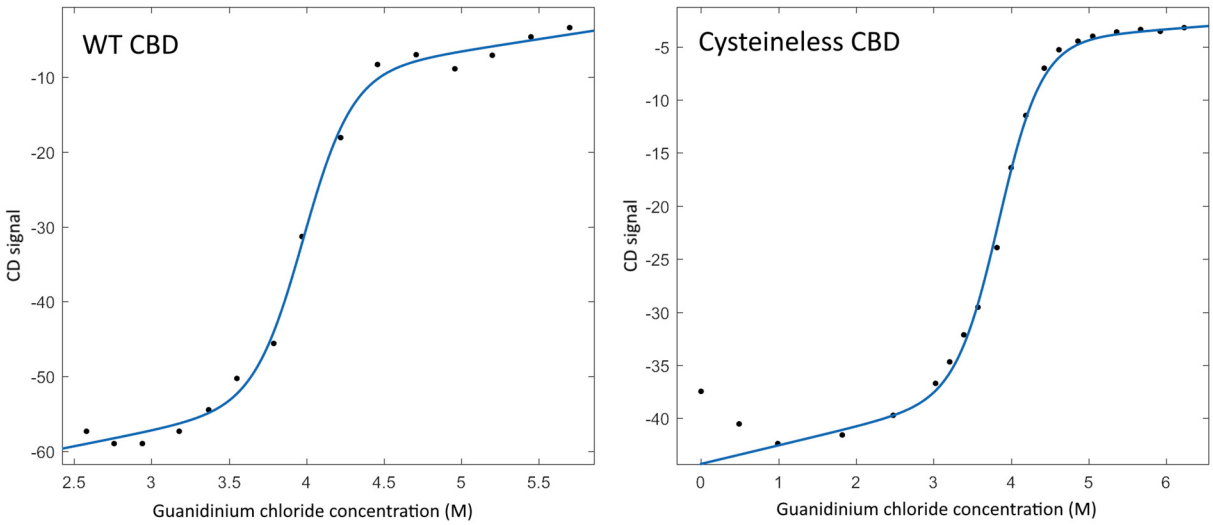


Figure 4.4 The unfolding profile of WT and cysteineless CBD measured by their CD signal at increasing guanidinium concentrations. Curves were fit by non-linear regression. WT CBD unfolding profiles monitored by fluorescence and circular dichroism collected by Laura Rosen and cysteineless CBD unfolding profiles monitored by circular dichroism collected by Matthew Draelos.

^{1}H-^{15}N HSQC NMR spectra of the WT and cysteineless CBD were compared to determine if the lack of cysteines changed the structure of the CBD. Nearly all peaks in the NMR spectrum of the cysteineless CBD overlay with those of the WT CBD, except for a small number of peaks corresponding to residues near the site of the mutations (Figure 4.5). This indicates that the structure of the CBD is not affected by the replacement of its native cysteines with alanines. Some peaks in the cysteineless CBD NMR spectrum are slightly broader than their WT counterparts which may be due to the increased local fluctuations of the cysteineless CBD reflecting the destabilization observed in the denaturation melts.

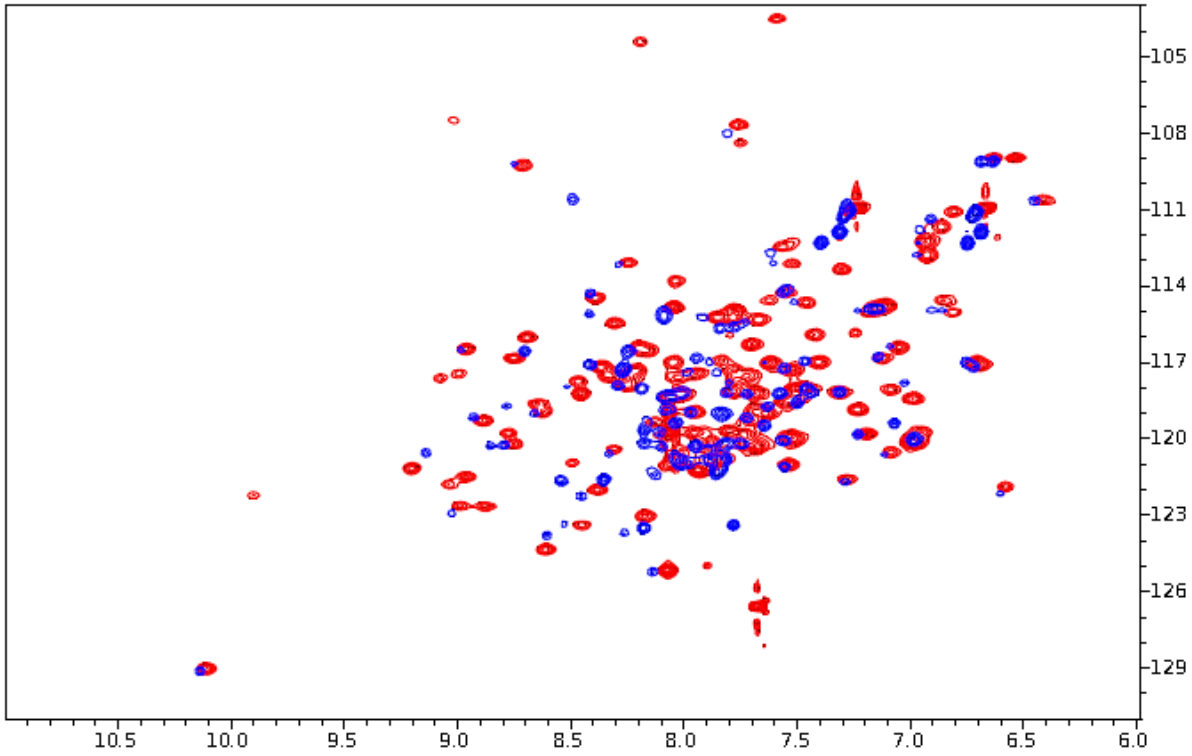


Figure 4.5 ^{1}H-^{15}N HSQC at 303K showing WT σ^{54} CBD (69-198) in red (Hong, Doucleff, and Wemmer 2009) overlayed with σ^{54} CBD (69-198) with N- and C-terminal cysteines, but otherwise cysteineless, in blue. The high number of overlapping peaks suggests that the cysteine to alanine mutations do not result in any major structural changes.

To determine if the cysteine to alanine mutations affected the local structure and stability of any of the helices, hydrogen exchange experiments were performed. Individual peak intensities of the assigned peaks in a 1 hour ^{1}H-^{15}N HSQC spectrum of CBD were monitored over time as they exchanged with the D_2O solvent. We observed similar timescales for the disappearance of amide peaks due to exchange in both the WT and cysteineless CBD (Figure 4.6). Residues in both the four-helix bundle (e.g. V127) and three-helix bundle (e.g. I155) show similar exchange rates to their WT counterparts

indicating that both of the two subdomains remain intact in the cysteineless CBD. This confirms that the cysteineless CBD is a suitable construct to study the hydrophobic interface between its two subdomains to determine if it could serve as a conformational fracture point when force is applied to it. In both WT and cysteineless CBD, we observed faster exchange rates for hydrophobic core residues on the seventh helix than those on the other six helices, perhaps indicating that the seventh helix is less stable than the others.

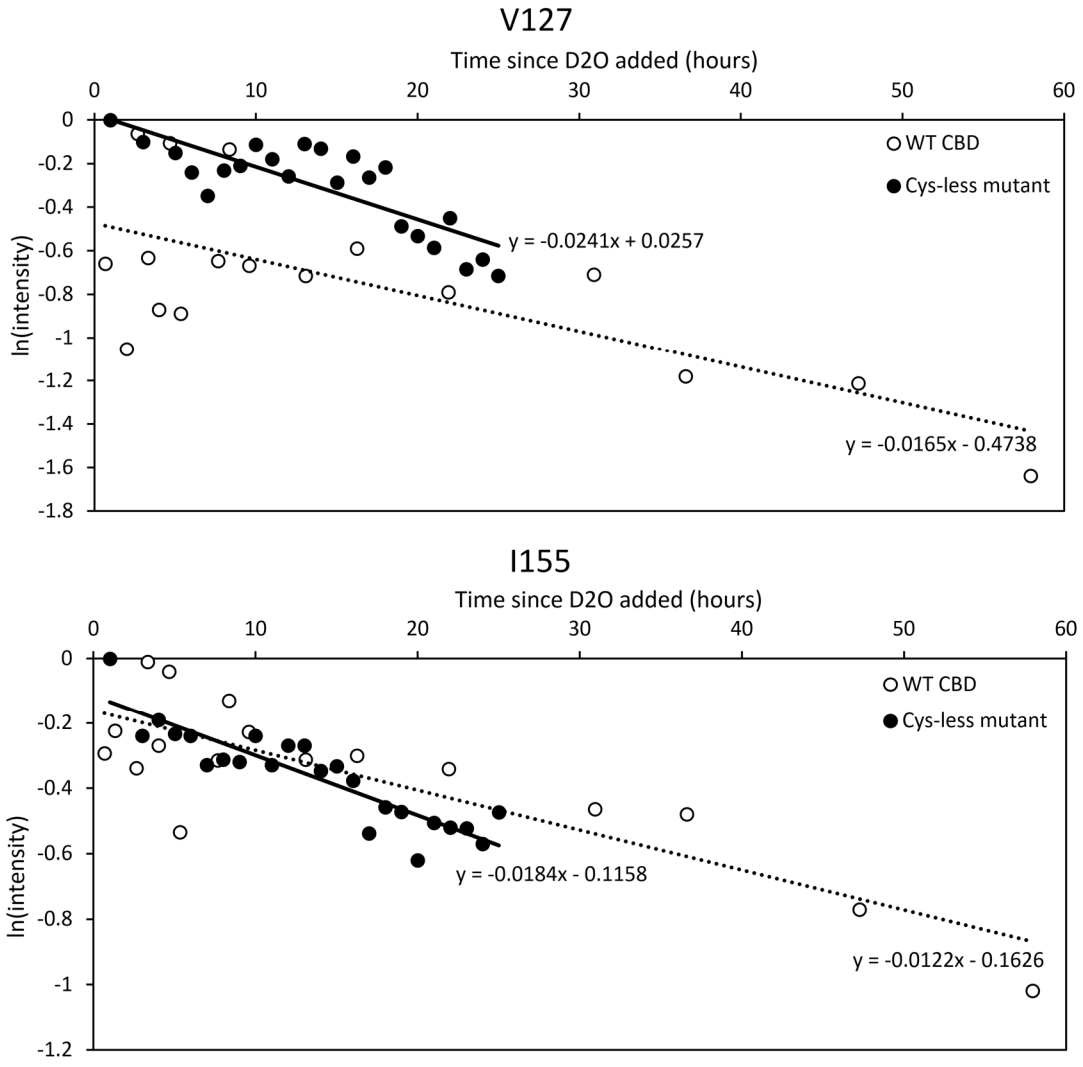


Figure 4.6 Hydrogen exchange peak intensities were monitored by 1 hour ^{1}H-^{15}N HSQCs for 1-3 days. Intensities of two amino acids in the hydrophobic core, V127 (top) and I155 (bottom), found on the four-helix bundle and three-helix bundle respectively, are shown for both the WT (white) and cysteineless (black) σ^{54} CBD. Comparable exchange rates between the WT and cysteineless CBD residues are observed for V127 and I155 as well as other residues in the hydrophobic core (not shown). The similar exchange rates of these residues indicate that the local stability of the four- and three-helix bundles are not affected by the cysteine to alanine mutations or the addition of N- and C-terminal cysteines. WT CBD NMR spectra collected by Laura Rosen and cysteineless CBD NMR spectra collected by Matthew Draelos.

A cysteineless CBD truncation mutant (69-176) without the seventh helix (the final helix of the three-helix bundle) was also purified. Many peaks in the ^{1}H-^{15}N HSQC NMR spectrum of the truncation, in particular those further away from the three-helix bundle, also overlap with the peaks of the WT and cysteineless CBD (Figure 4.7) indicating that it retains some of the same structure. However, the peaks are also significantly broader in the CBD lacking the seventh helix than they are in the WT and cysteineless CBD, indicating that the removal of the seventh helix significantly destabilizes the structure and allows more conformational fluctuations.

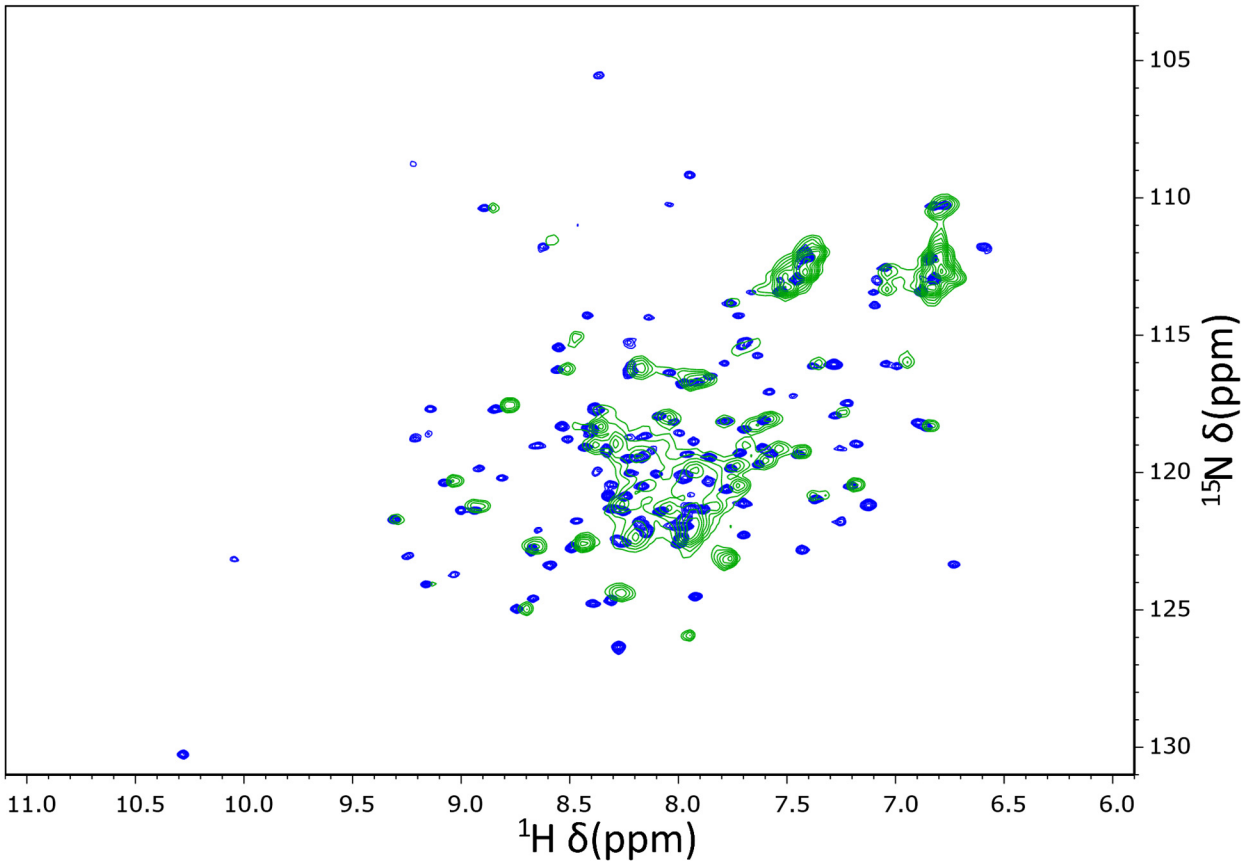
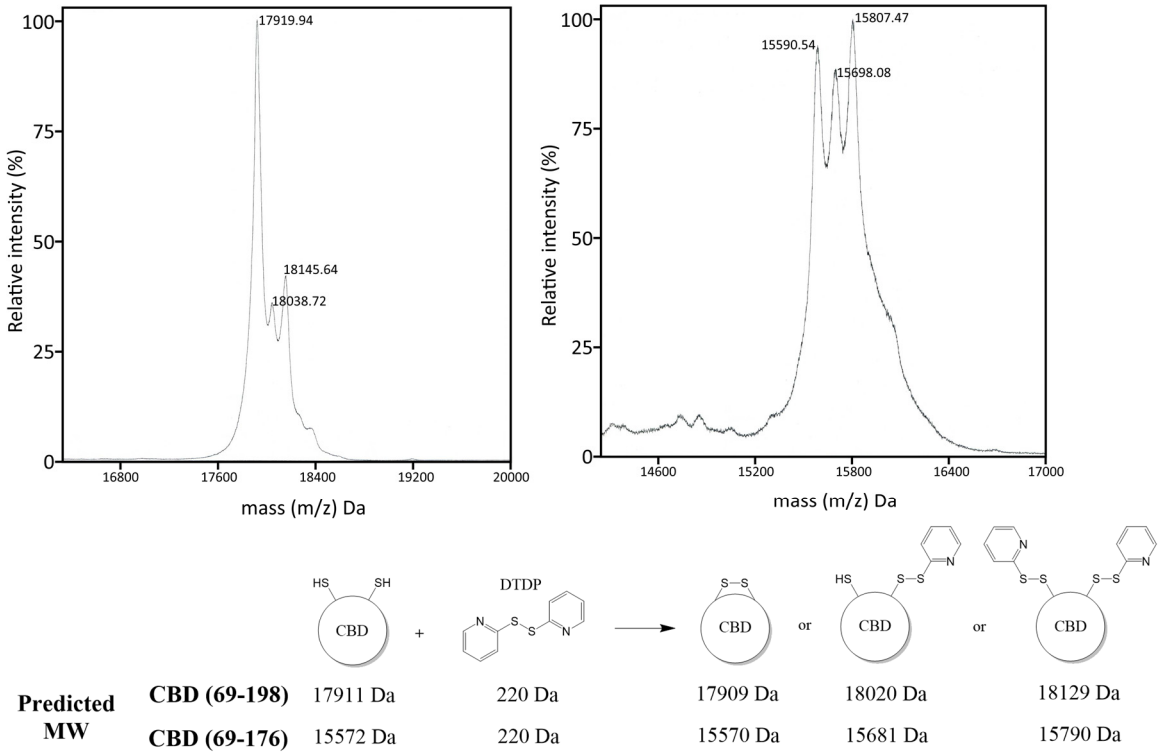


Figure 4.7 ^{1}H-^{15}N HSQC at 303K showing the cysteineless σ^{54} CBD (69-198) in blue (Hong, Doucleff, and Wemmer 2009) and cysteineless CBDΔhelix7 (69-177) in green. The truncation of the seventh helix results in significant broadening but few peak shifts, suggesting it folds into the same structure but is destabilized.

4.3.2 Successful derivatization

The N- and C-terminal cysteines of *A.a.* σ^{54} CBD (69-198) and the truncated CBD (69-176) were derivatized with dithiodipyridine (DTDP). To avoid the formation of a disulfide bond between the N- and C-terminal cysteines, the CBD was denatured in 6 M guanidinium chloride before reacting with DTDP and refolding back into its native

structure. The successful derivatization reaction between the cysteines and DTDP was confirmed by MALDI-TOF mass spectrometry (Figure 4.8).



Predicted MW	**CBD (69-198)**	17911 Da	220 Da	17909 Da	18020 Da	18129 Da
	CBD (69-176)	15572 Da	220 Da	15570 Da	15681 Da	15790 Da

Figure 4.8 Top: Mass spectra showing three peaks of varying intensity for the CBD (left) and CBDΔhelix7 (right) after activation with DTDP but before reacting with free thiols on the DNA handles. Below the expected masses are shown. The derivatization of zero, one or two DTDP groups were observed.

The DTDP-activated core binding domain constructs were reacted with DNA strands containing a thiol on one end and either biotin or digoxigenin on the other. The protein-DNA handle hybrids were run on a native gel to verify that a small fraction of the CBD protein construct was successfully labeled with two DNA handles (Figure 4.9). Of these, only half will have digoxigenin attached to one DNA handle and biotin attached to the other, but only these CBD molecules with two different handles will be caught by both beads in the optical trap (Cecconi et al. 2008).

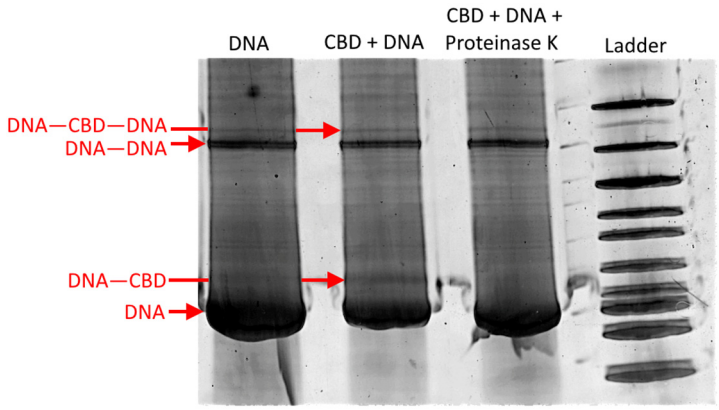


Figure 4.9 Native gel showing a successful handle attachment reaction. Lane 1 is the purified, deprotected DNA handle alone, some of which forms a disulfide with a second handle. Lane 2 is the DNA handle successfully reacted with DTDP activated CBD, which shows a second band slightly above the single DNA handle with one protein attached to one DNA. The DNA-CBD-DNA band is usually too faint to see due to the rarity of two successful reactions occurring on the same molecule, but likely exists in low quantities when a band showing one DNA bound to the protein is observed. Lane 3 shows the product from lane 2 after digesting with Proteinase K, which proteolyzes the CBD and leaves only the DNA handles. Lane 4 is a DNA ladder.

Core binding domain molecules with a digoxigenin handle on one end and a biotin handle on the other were bound to polystyrene beads coated with anti-digoxigenin antibodies or streptavidin, respectively. To achieve this, CBD protein molecules tethered with DNA handles to beads on either end were flowed through a chamber where the beads were captured by a pipette on one end and a laser on the other. Once a molecule was caught in the optical trap, we performed either force ramp or force clamp experiments to identify any short-lived intermediates along the unfolding pathway brought about by the application of force to the N- and C-termini of the core binding domain. In particular, we were interested to see if force alone could drive the unfolding of one of the two CBD subdomains, which would provide some evidence of a conformational fracture point.

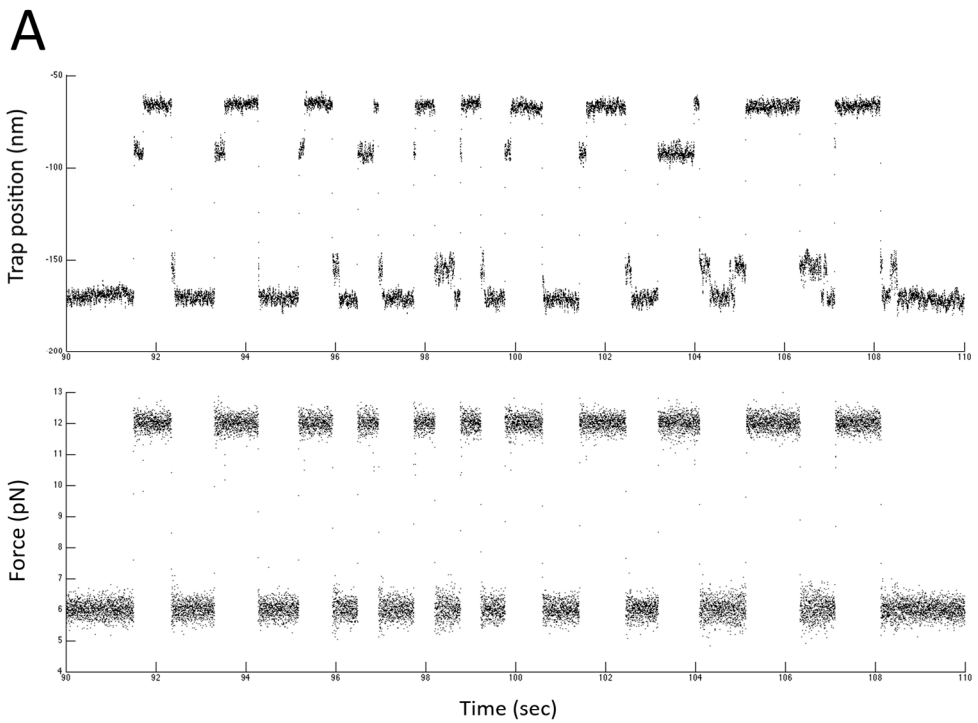
4.3.3 Force clamp experiments

The tethered core binding domain was subjected to force clamp experiments using optical tweezers (Greenleaf et al. 2005). The CBD was held alternately at a constant high force (in the 9-14 pN range) or a constant low force (in the 4-6 pN range) and the extension of the molecule was observed over time by monitoring the position of the beads. When the protein is folded its length is minimized and the extension of the beads will be low, and conversely when the protein unfolds its length is maximized and the extension of the beads will be high. Any small amount of time that the protein spends as a partially unfolded intermediate, e.g. if the three-helix bundle unfolds before the

four-helix bundle at high force, will result in extensions in between the folded and unfolded states. By alternating between high and low force to unfold and refold the protein, we can determine the amount of time that a single CBD molecule spends at each extension, and consequently the amount of time it spends folded, unfolded, or partially folded. With enough data and sufficiently high force, these occupancies will form Gaussian distributions. Sparsely populated partially unfolded intermediates may be hard to identify at low force but become more prominent at high force (Cecconi et al. 2005). In particular, we looked for a peak at an extension roughly halfway between the extension of the folded and unfolded states that might represent the unfolding of just the three-helix bundle.

In a typical force clamp experiment the force is suddenly raised to unfold the protein and is held constant for a short time, then suddenly lowered to refold the protein and held constant again before the same process is repeated. A sample of two minutes of force clamp data shows the CBD unfolding rapidly shortly after the force is raised to 12 pN and refolding rapidly shortly after the force is lowered to 6 pN (Figure 4.10A). By counting the amount of time the CBD molecule spends at each extension across both forces, a histogram of the occupancy of the folded and unfolded states at high (e.g. 12 pN) and low (e.g. 6 pN) force can be calculated (Figure 4.10B). From this histogram, four Gaussian distributions are visible, they are from the lowest to highest extensions (left to right): folded CBD at low force, unfolded CBD at low force, folded CBD at high force, and unfolded CBD at high force. Another Gaussian distribution is less obvious but appears as a small shoulder on the high force folded state distribution and represents a partially unfolded intermediate of the CBD.

Sample force clamp data at 6 pN and 12 pN



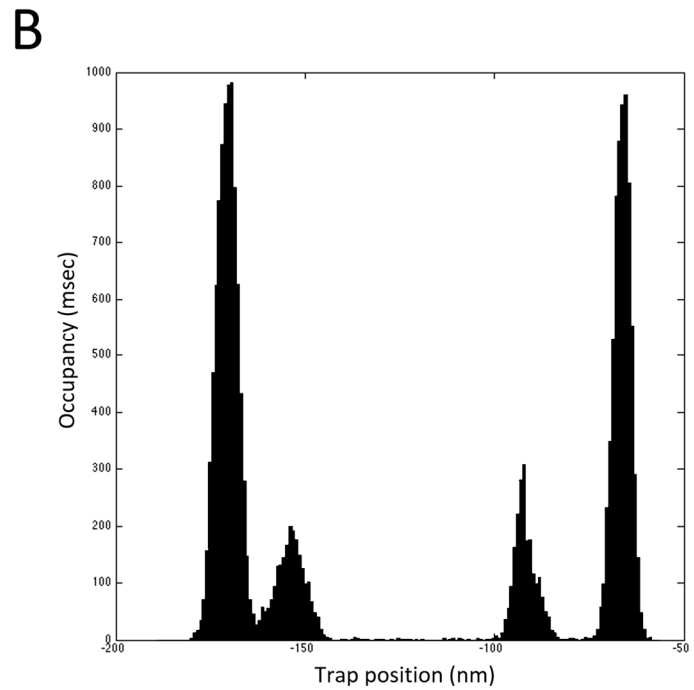


Figure 4.10 (A) A sample two minutes of data showing the trap position (i.e. the extension of the molecule) over time while CBD is held at 6 pN or 12 pN. (B) A histogram counting the amount of time the CBD spends at each extension during the 6 to 12 pN force clamp experiments. From left to right the peaks at these extensions correspond to the folded CBD at 6 pN force, unfolded CBD at 6 pN, folded CBD at 12 pN, and unfolded CBD at 12 pN. A shoulder on the right side of the folded CBD at 12 pN represents a partially unfolded intermediate which has a length ≈ 4 nm greater than the folded state.

Force clamp experiments were used to calculate histograms of the occupancy of the CBD at each extension for forces between 5 and 14 pN (Figure 4.11). The histograms were fit to a function with three Gaussians representing the unfolded, partially unfolded intermediate, and folded states. As the force is increased, the distance between the folded and unfolded states also increases going from 8 nm at 5 pN to 28 nm at 14 pN. A small shoulder caused by a partially unfolded intermediate appears on the distribution of the folded state and grows more prominent at higher forces (Figure 4.12). The average difference in extension between the folded and partially unfolded state found at high force (between 9 and 14 pN) is a distance of 4±1.6 nm.

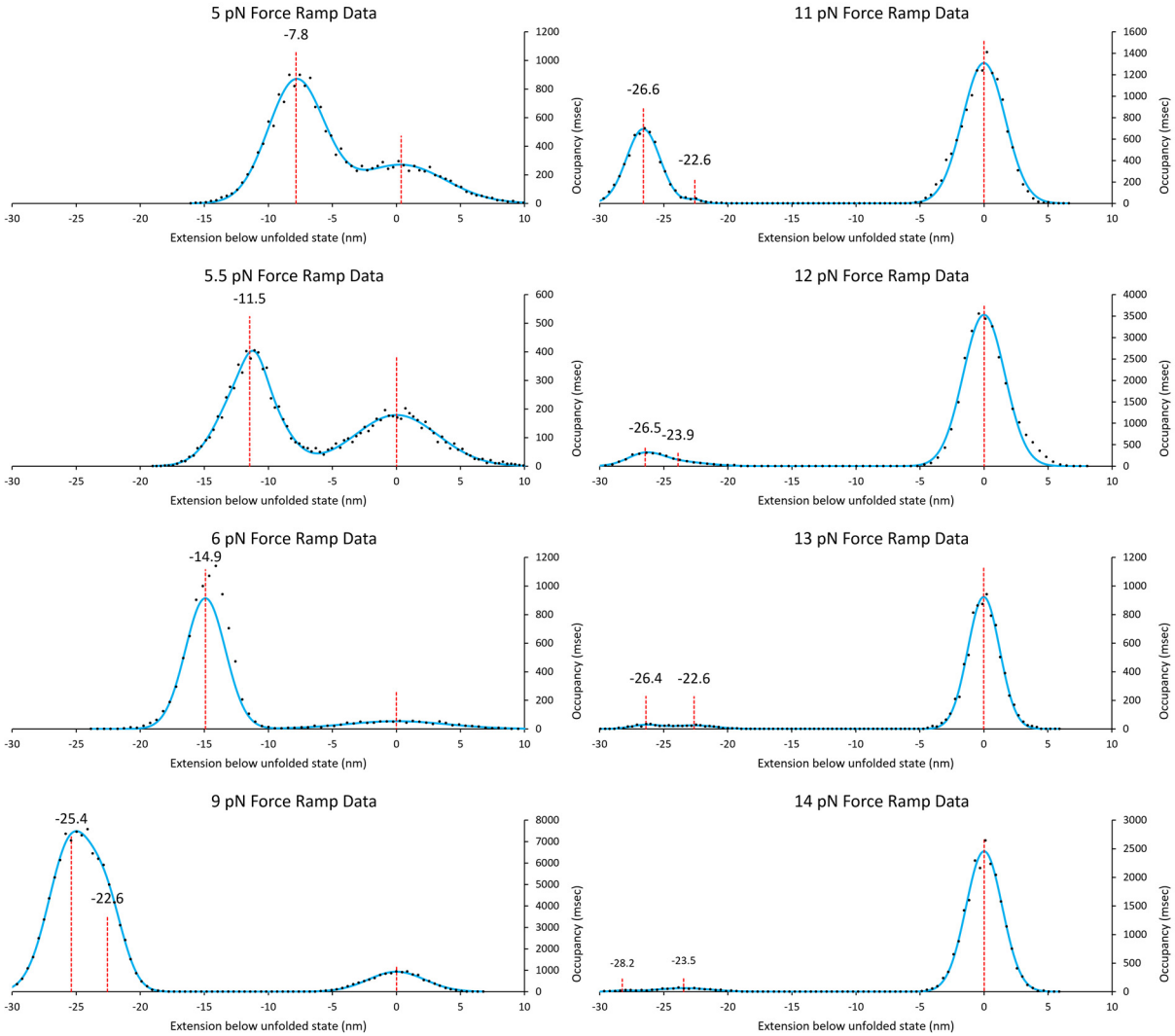


Figure 4.11 The data points represent the amount of time (occupancy) in milliseconds that the CBD spends at each extension in nm below the unfolded state at eight forces between 5 and 14 pN. These histograms are fit to a function that is the sum of three Gaussian distributions (blue). The center of each Gaussian distribution is marked (dotted red line) and represents the difference in extension between the folded or partially unfolded state and the unfolded state. No partially unfolded state was detected at low forces (5-6 pN).

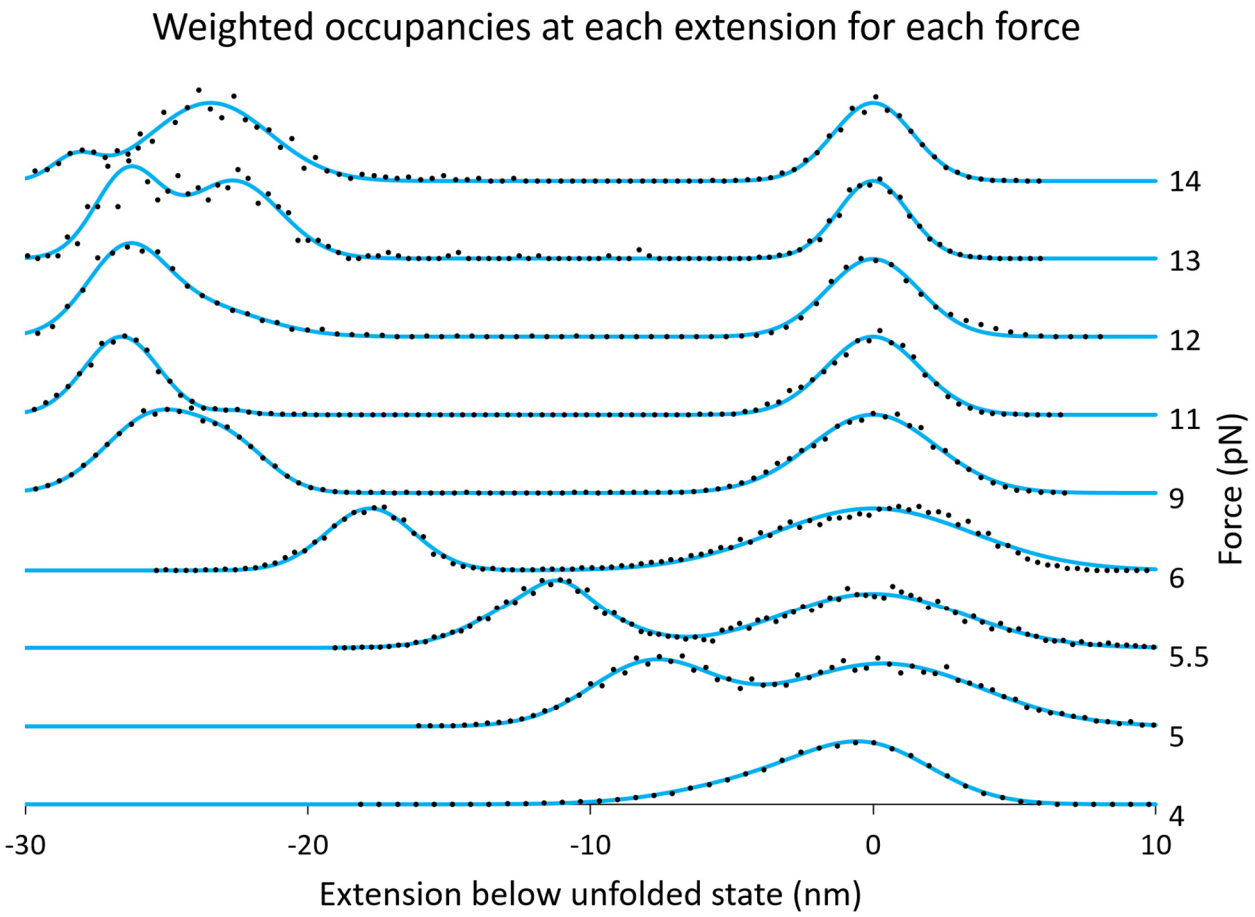


Figure 4.12 The histogram data of extensions below the unfolded state at each force is shown but with the unfolded and folded states separately weighted to reach the same height. This allows for easier visualization of the distances between the peaks. The extensions in nm of the folded state state gradually increase at higher forces. The partially folded state is visible at forces above 9 pN and becomes especially prominent at 13 – 14 pN.

4.3.4 Force ramp experiments

Force ramp experiments were used to apply increasing force to a CBD molecule by gradually separating the beads. During each trace, the force was increased from 3 to 15 pN and then back to 3 pN while the extension of the molecule was monitored by recording the position of the bead. At high enough force, the CBD cooperatively unfolds resulting in a sudden decrease in the force seen as a rip in the trace. When the beads are brought close together the CBD refolds resulting in a sudden increase in the force. This process was repeated 27 times on 4 separate CBD molecules (Figure 4.13A). Overlaying the force ramps highlights the forces at which the unfolding and refolding transitions occur during the many trials (Figure 4.13B). Plotting a histogram of the number of

unfolding or refolding events that occur at each force (Figure 4.13C), we determined the average refolding force was 5.6±0.3 pN and the average unfolding force was 12±1 pN.

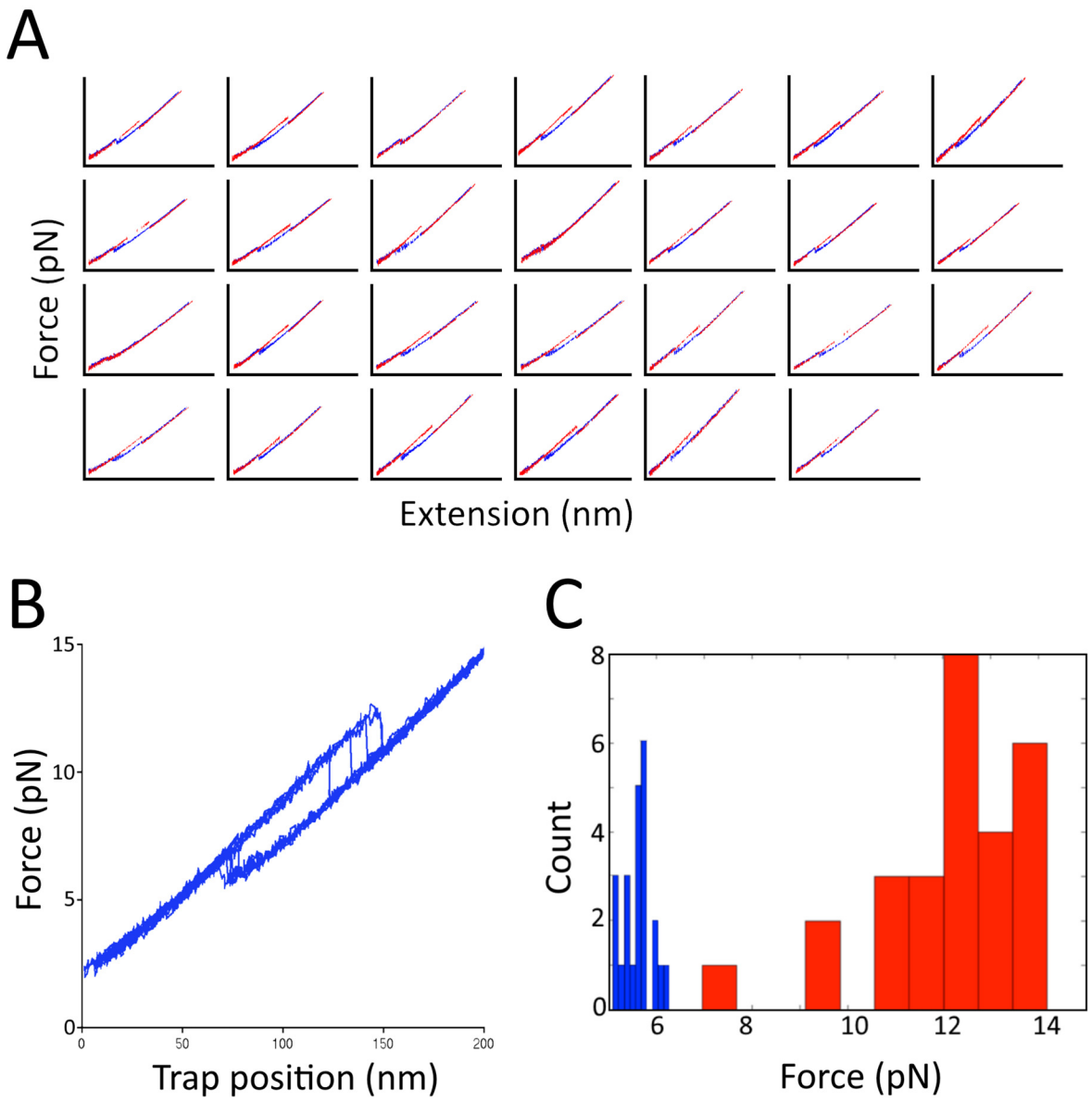


Figure 4.13 (A) 27 force ramp traces performed on 4 different CBD molecules. The gradually increasing extension (red) results in an unfolding event and sudden decrease in the force. The gradually decreasing extension (blue) results in a refolding event and sudden increase in force. (B) Overlaying all 27 traces reveals the distribution of the forces required for unfolding and refolding the CBD. (C) A histogram showing the distribution of the forces that caused CBD to unfold (red) or refold (blue), with an average unfolding force of 12±1 pN and an average refolding force of 5.6±0.3 pN.

To get a more accurate distribution of unfolding and refolding forces to estimate the free energy of folding, a significantly larger number of force ramp trials need to be

performed (Hummer and Szabo 2001). However, even in the preliminary force ramp data we do not see any deviations from two-state unfolding. If one of the CBD subdomains remains folded after the other unfolds we would expect to see two rips in each trace for the unfolding of each of the two subdomains at two different forces. Instead, the two subdomains unfold cooperatively, though we can't rule out the possibility that their unfolding behavior might change when in complex with RNA polymerase.

4.4 Discussion

4.4.1 Unfolding of the seventh helix of the core binding domain

A remaining puzzle in the mechanism of σ^{54} transcription initiation is how the action of transcriptional activators on the N-terminal activator interacting domain (AID) can lead to conformational changes in the C-terminal DNA binding domains that melt the DNA. The structure of the core binding domain (CBD), which lies between the site of activation and the site of DNA melting in the protein sequence, revealed two distinct subdomains held together by a hydrophobic interface (Hong, Doucleff, and Wemmer 2009). One subdomain, a four-helix bundle, remains folded in its native structure in the absence of the other subdomain, a three-helix bundle. This suggested a mechanism where the application of force on the N-terminal side of the CBD could selectively unfold the three-helix bundle, causing large scale conformational rearrangements to the DNA binding domains on the C-terminal side of the CBD that would lead to the formation of the open complex.

To study the validity of this mechanism, we performed single molecule optical tweezer experiments to monitor the folding and unfolding of the CBD at different forces. In all force ramp experiments where the CBD was subjected to gradually increasing force, the CBD molecules exhibited only a single unfolding or refolding event. Though the four-helix bundle is able to fold in the absence of the three-helix bundle, the hydrophobic interface between the subdomains stabilizes the three-helix bundle enough that it can withstand forces in excess of the ≈ 12 pN required to unfold the entire CBD. The cooperative unfolding of both subdomains casts doubt on the hypothesis that they act as a fracture point in the σ^{54} structure.

We also performed force clamp experiments on CBD where force is held constant but alternates between a high force that unfolds the protein and low force that refolds it. In

these experiments at forces between 4 – 14 pN, the CBD was never detected occupying a partially unfolded intermediate attributable to the unfolding of just the three-helix bundle. However, at high forces we did observe a small shoulder on the Gaussian distribution of the folded state corresponding to an additional extension of 4±1.6 nm beyond the length of the folded state. This sparsely populated intermediate along the unfolding pathway grows in occupancy relative to the folded state at higher forces. If the 10 amino acid seventh helix unfolds independently, it would create an additional extension of 3.65 nm using the estimate of 0.365 nm per unfolded amino acid (Dietz and Rief 2004). Based on the structure of the CBD, there would also be a ≈ 1.4 nm increase in the distance between the two points of force application on the N- and C-terminus of the folded structure once the seventh helix unfolds. Together this would result in a total additional extension of ≈ 5 nm, roughly in line with the 4±1.6 nm additional extension of the observed intermediate. NMR hydrogen exchange data also show that buried residues attached to this seventh helix have faster exchange rates than buried residues on the other helices, suggesting the seventh helix is the least stable region of the CBD. Together this suggests that the partially unfolded intermediate is the unfolding of the seventh helix of the CBD.

To verify that this partial unfolding intermediate is actually the unfolding of the seventh helix, a cysteineless CBD with the seventh helix truncated (69-176) was purified. Though the peaks in its NMR spectrum overlap well with those in the full CBD indicating no major structural changes, they are also significantly broadened due to the reduced stability of the three-helix bundle. Optical tweezer experiments on the truncated version have proven difficult, and we have yet to verify that removal of the seventh helix will eliminate the partially unfolded intermediate observed in the experiments on the full CBD. In all optical tweezers experiments, we found no evidence of separate unfolding events for the two CBD subdomains to support the hypothesis that CBD serves as a conformational fracture point in σ^{54}. However, we did not study the effect of force on the CBD subdomains in their native complex with RNA polymerase, which may produce different results than when force is applied to CBD alone.

4.4.2 Future directions

We found no evidence from our optical tweezers experiments on the CBD alone that the four-helix bundle subdomain remains folded after the three-helix bundle has unfolded. We therefore cannot support the mechanism that the CBD functions as a force-dependent fracture point in the σ^{54} structure. However, it is possible that

force-dependent conformational rearrangements of the CBD require it to be in its native complex with RNA polymerase. To test this, we can apply force to the cysteineless CBD in the presence of core RNAP, though it's unclear if the structure of CBD alone bound to the core RNA polymerase will be sufficiently similar to that of full length σ^{54} found in the RNAP holoenzyme. Alternatively, a cysteineless *E.c.* $\sigma^{54}\Delta$AID, lacking only the N-terminal AID, can be used. This presents some technical challenges in removing cysteines from σ^{54} and attaching DNA handles in appropriate places without disrupting the σ^{54} structure or its ability to form the native complex with RNA polymerase. With handles placed immediately before and after the CBD we could apply a force to the CBD that mimics the force that the transcriptional activators may apply to the N-terminal AID in the threading mechanism (see Chapters 2 and 3). If this force generates a conformational rearrangement in the CBD before it unfolds, we would notice a partially unfolded intermediate between the folded and unfolded state.

A more challenging experiment, but one that could provide the significant insight into the activation mechanism of σ^{54}, would be to reconstitute the full σ^{54} system in the optical tweezers, including the $\sigma^{54}\Delta$AID, core RNAP, and the promoter DNA. By attaching one handle to the N-terminal end of the CBD and another handle to the promoter DNA itself, we could test the ability of σ^{54} to melt DNA with force alone. Using fluorescent probes attached near the -12 promoter element and site of DNA opening, we could observe the moment that the RNAP holoenzyme melts the DNA. With a combined single-molecule fluorescence and optical tweezers setup, we could simultaneously measure the force vs. extension and fluorescence to determine if the application of force alone can drive conformational changes in the σ^{54}-RNAP holoenzyme that cause it to open DNA. If successful, this would provide important evidence to support the force dependent mechanism of σ^{54} transcription initiation.

4.5 Materials and Methods

4.5.1 DNA Handle Generation

DNA handles were generated by performing large scale PCR on the pGMEX-1 plasmid with 5' thiol-GCT-ACC-GTA-ATT-GAG-ACC-AC as the forward primer and either 5' digoxigenin- or 5' biotin-CAA-AAA-ACC-CCT-CAA-GAC-CC as the reverse primer. 10 mL reaction volumes containing 1x Taq buffer, 1.5 mM magnesium chloride, 0.2 mM dNTPs, 0.5 μM of each primer, 0.04 ng of pGMEX-1 plasmid, 16.2 mM DTT, and 0.025 U Taq polymerase were loaded into GeneMate 96-well PCR plates sealed and loaded

into a thermocycler. Following the reaction, DNA handles were purified using a QIAGEN HiSpeed Plasmid Midiprep kit.

4.5.2 Preparation of σ^{54} core binding domain

Aquifex aeolicus σ^{54} residues 69-198 were cloned into pET-28a vectors behind an N-terminal His_6 tag. Cysteines at sites 119 and 140 were replaced with alanines and the tyrosine at the N-terminal residue 69 and lysine at residue 198 were replaced with cysteines by QuikChange site directed mutagenesis. The plasmid was transformed into BL21 Rosetta cells, grown in LB, induced at $A_{600nm} = 0.6$ with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and grown for three more hours at 37°C. Cells were pelleted and resuspended in 20 mM sodium phosphate, 0.5 M NaCl, and 2 mM β-mercaptoethanol at pH 7.4, and sonicated. Lysates were heated to 75°C for 20 minutes to denature non-thermophilic proteins and pelleted at 30,000 rpm. The supernatant was incubated with Ni-NTA resin overnight, then washed with 20 mM sodium phosphate, 0.5 M NaCl, 2 mM β-mercaptoethanol and 20 mM imidazole at pH 7.4 and eluted with 20 mM sodium phosphate, 0.5 M NaCl, 2 mM β-mercaptoethanol and 300 mM imidazole at pH 7.4.

4.5.3 DTDP Activation

Purified σ^{54} CBD was reconcentrated into 6 M guanidinium chloride, 100 mM sodium phosphate, pH 7.0 in a Millipore spin concentrator. Denatured CBD was reduced for 1 hour with 10 mM dithiothreitol (DTT) in the presence of excess dithiodipyridine (DTDP). Reduced CBD was buffer-exchanged by gravity filtration on a sephadex G-15 column into 0.1 mM sodium phosphate, pH 5.5 with 25 molar excess DTDP and no DTT and allowed to react with DTDP overnight at room temperature. DTDP attachment to the N- and C-terminal cysteines was verified by MALDI-TOF mass spectrometry on an Applied Biosystems Voyager System 6322 mass spectrometer.

4.5.4 Attachment of the Handles

160 nmoles of the digoxigenin and biotin DNA handles in a 1:1 ratio were mixed and concentrated to 100 μL. The free thiol end of the handles was reduced by adding 40 molar excess DTT for 1 hour at room temperature before buffer exchanging the handles out of DTT and into 100 mM sodium phosphate, pH 8.0 using Bio-Rad Biospin 6 Chromatography columns. 80 nmoles of DTDP-activated CBD was reacted with the DNA handle mixture along with 0.05% tween. The reaction was left overnight at room

temperature and successful attachment of two DNA handles was verified by visualizing the protein-DNA hybrids on a native gel. Only half of the protein with two successfully attached DNA handles is expected to have one digoxigenin and one biotin.

4.5.5 Molecular Tweezers experiments

1 μL of protein, containing some CBD molecules covalently attached to exactly one digoxigenin and one biotin labeled DNA handle, was reacted with polystyrene beads coated with anti-digoxigenin antibodies for 10 minutes at room temperature to attach the digoxigenin handle end of the CBD to the bead. These reactions were flowed into the optical tweezer trap and the beads were caught with a laser. By moving the position of the laser, the anti-digoxigenin beads were brought into close proximity of a second polystyrene bead coated in streptavidin and held by a pipette. When a biotin DNA handle on one side of CBD comes close enough to the streptavidin coated bead on the pipette they bind together to form a tether. These tethered CBD molecules were subjected to two kinds of experiments. Force clamp experiments apply rapid changes in force to the protein to measure the occupancy of the unfolded, intermediate and native states. Force ramp experiments apply gradual increases in force to determine a distribution of the average force required to unfold the native state into either an intermediate or fully unfolded state and then refold it back again. The optical tweezer experiments were performed with the help of Jesse Dill on a single bead optical trap assembled by the Marqusee lab, as described in (Shank et al. 2010). The optical tweezers sample at 1 kHz for a time resolution of 1 millisecond, have an optical-trap spring constant of 0.075 pN/nm which corresponds to a spatial resolution of roughly 7.4 nm and force resolution of 0.55 pN. Optical tweezer data was analyzed using OPTIMAL: **OP**tical **T**weezers **I**ntegrated **M**atlab **A**nalysis **L**ab (Dill 2012).

4.5.6 NMR and Hydrogen Exchange

^{15}N-labeled cysteineless σ^{54} CBD (69-198) and cysteineless σ^{54} CBDΔhelix7 (69-176) were expressed in M9 minimal media with $^{15}NH_4Cl$ as the only nitrogen source. ^{1}H-^{15}N HSQC spectra of purified protein in H_2O + 10% D_2O NMR buffer (25 mM sodium phosphate, 100 mM NaCl, 1 mM dithiothreitol, pH 7.0) were obtained at 298K on a Bruker DRX 800-MHz spectrometer. For hydrogen exchange experiments, ^{15}N-labeled protein was lyophilized and mixed with 100% D_2O NMR Buffer immediately before taking a series of 1 hour ^{1}H-^{15}N HSQCs for 1-3 days to determine amide exchange rates, particularly for residues in the hydrophobic core. Exchange rates of residues on the cysteineless mutant and WT σ^{54} CBD (69-198) were compared to determine the impact

that the cysteine to alanine mutations have on the overall stability of the structure and the individual subdomains.

4.5.7 Circular dichroism stability measurements

Purified protein was equilibrated in increasing concentrations of 0 – 6 M guanidinium chloride for 36 hours at 25°C. Guanidinium concentration was corrected by measuring the refractive index on a refractometer. Circular dichroism signal was measured on an AVIV circular dichroism spectrometer at a wavelength of 222 nm. The circular dichroism signal was measured at 1 second increments over a 60 second interval and averaged. Data were fit to a sigmoidal curve using nonlinear regression to calculate the m-value and free energy of unfolding ($\Delta G_{unfolding}$).

Chapter 5: Activation of the piezophilic *Shewanella violacea* NtrC receiver domain

5.1 Summary

In this chapter, I describe work examining the activation of the NtrC regulatory receiver domain from the piezophilic bacteria, *Shewanella violacea*. In response to high pressure, *S. violacea* expresses genes regulated by σ^{54} and the transcriptional activator, NtrC. Previous work characterizing the *Escherichia coli* NtrC receiver domain revealed an allosteric mechanism where phosphorylation of a conserved aspartate changes the relative energies of the active and inactive states and results in a majority of receiver domains adopting the active conformation. We hypothesized that high pressure might drive a similar change in the equilibrium populations of the active and inactive *S.v.* NtrC receiver domain, thereby bypassing the other components of the NtrC signaling pathway and immediately initiating the σ^{54}-dependent expression of pressure response genes. We used high pressure NMR spectroscopy to study the structural changes to the receiver domain but found no evidence that the relative population of the active state increases at higher pressures. We conclude that a different component in the *S.v.* NtrC signaling pathway is responsible for sensing pressure and triggering σ^{54}-dependent gene expression in response.

5.2 Introduction

5.2.1 The regulatory receiver domain of NtrC

Transcriptional activators are themselves activated by a variety of different N-terminal regulatory domains (Studholme and Dixon 2003). Once these regulatory domains are activated, they promote the oligomerization of the central AAA+ ATPase domain, which interacts with σ^{54} to drive open complex formation (De Carlo et al. 2006). The regulatory domain of NtrC and other related transcriptional activators is known as a receiver domain, and is the response regulator of a two component signal transduction pathway leading to transcription of genes responding primarily to limited nitrogen

(Reitzer 2003). The receiver domain of *E. coli* NtrC is phosphorylated and activated by the histidine kinase NtrB. NtrB is autophosphorylated when it senses low levels of nitrogen in the environment, and that phosphate is transferred to the NtrC receiver domain (Keener and Kustu 1988). Once the receiver domain is phosphorylated, NtrC oligomerizes into the active conformation which leads to σ^{54} activation and the initiation of transcription at genes necessary to respond to nitrogen limiting conditions (Ninfa, Reitzer, and Magasanik 1987).

5.2.2 Two-component signal transduction of σ^{54}-mediated glnAp2 transcription initiation

Signal transduction is any process where a protein responds to an environmental stimuli leading to transcription of genes necessary for the cell's proper response to the environmental changes. Two component signal transduction typically involves two proteins: a histidine kinase which transfers the signal and a response regulator which receives the signal and regulates the cell's response, often by acting as a transcription factor (Ulrich, Koonin, and Zhulin 2005). For NtrC-mediated transcription initiation, the histidine kinase NtrB acts as a phosphodonor and transfers a phosphate group from His139 on NtrB to Asp54 on the response regulator NtrC (Ninfa and Bennett 1991)(Weiss and Magasanik 1988). In addition to acting as a kinase and phosphodonor to NtrC, under certain circumstances, NtrB can also act as a phosphatase that dephosphorylates and thus deactivates NtrC (Keener and Kustu 1988). The rate of the phosphodonor and phosphatase reactions of NtrB determines whether NtrC is predominantly phosphorylated and active or dephosphorylated and inactive.

Whether NtrB acts as a phosphodonor to activate NtrC or a phosphatase to deactivate it is determined by a third protein, P_{II}. Under nitrogen limiting conditions, uridylyl transfer/uridylyl removing enzyme (UT/UR) acts as a uridylyltransferase while under nitrogen excess it acts as a uridylyl-removing enzyme (Bourret, Borkovich, and Simon 1991). P_{II} is only active when it is not uridylylated, which occurs when there is an excess of nitrogen. Active P_{II} stimulates the phosphatase activity of NtrB, which dephosphorylates and inactivates NtrC before sufficient phosphorylated NtrC can build up and activate transcription (Keener and Kustu 1988). Under limited nitrogen conditions, P_{II} is uridylylated with UMP and inactive thus allowing large amounts of phosphorylated NtrC to build up and initiate transcription, for example at the glnAp2 promoter of the glnALG operon coding for glutamine synthetase, NtrB, and NtrC.

Glutamine synthetase fixes ammonia in the cell by converting glutamate to glutamine (Figure 5.1).

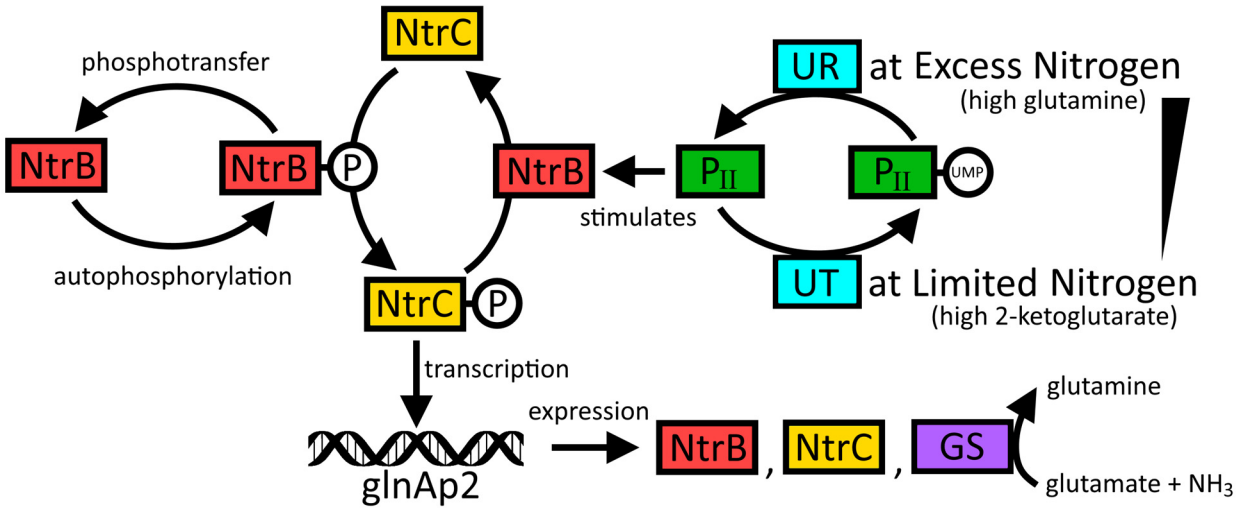


Figure 5.1 Overview of the NtrC two component signal transduction pathway. Proteins shown are the histidine kinase NtrB (red), the σ^{54} transcriptional activator NtrC (yellow), the nitrogen regulatory protein P_{II} (green) sometimes uridylylated with UMP, the uridylyltransferase at high nitrogen which is also a uridylyl removing enzyme at low nitrogen (blue), and glutamine synthetase (purple) the product of the glnA gene.

NtrC is a transcriptional activator of σ^{54}, that binds to enhancer sequences upstream of the glnA gene (Sasse-Dwight and Gralla 1988). The glnA gene has a σ^{70} (glnAp1) and a σ^{54} promoter sequence (glnAp2) but it is transcribed in much higher amounts by the σ^{54} pathway when NtrC is activated (Reitzer and Magasanik 1985). When its receiver domain is phosphorylated, NtrC hexamerizes into its active form and is able to initiate transcription of σ^{54} and transcribe the glnA gene coding for glutamine synthetase (Keener and Kustu 1988). Glutamine synthetase helps scavenge low levels of ammonia in order to incorporate nitrogen into the amino acid biosynthetic pathways necessary for cell survival (Tyler 1978). Once the cell begins to incorporate ammonia and the levels of glutamine have risen, UR removes the UMP from P_{II} leading to a dephosphorylation and deactivation of NtrC and reducing the levels of glnAp2 transcription (Bourret, Borkovich, and Simon 1991).

5.2.3 Mechanism of NtrC receiver domain activation

The *Salmonella enterica* NtrC receiver domain is activated by the histidine kinase, NtrB, which transfers the phosphate on its H139 to D54 on NtrC (Ninfa and Bennett 1991)(Weiss and Magasanik 1988). Phosphorylation of the NtrC receiver domain induces a conformational change that exposes a new hydrophobic surface (Kern et al.

1999). For *in vitro* experiments, this phosphorylation can be mimicked by addition of BeF_3^-, which associates with D54 in the same position as the phosphate (Hastings et al. 2003). In the activated state, the phosphorylated aspartate (or the beryllofluoride activated analog) stabilizes a reconfiguration of the side chains in the hydrophobic core between the fourth α-helix (residues 86-96) and the fifth β-strand (residues 100-101). In particular, Y101 on the fifth β-strand rotates in the activated state to fill a hydrophobic pocket (Figure 5.2). This causes α-helix four to undergo a significant positional rearrangement that changes its angle with respect to the fifth β-strand.

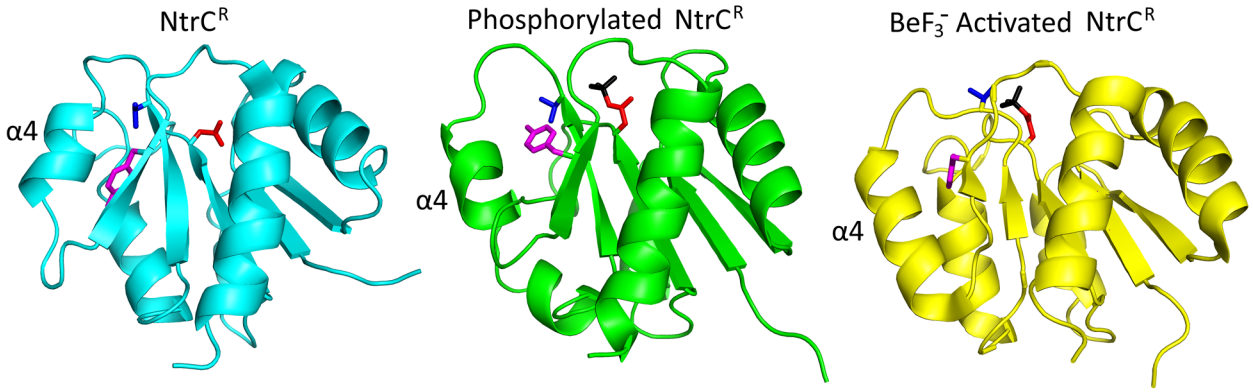


Figure 5.2 Ribbon representation of apo-$NtrC^R$ (left, 1DC7), phosphorylated $NtrC^R$ (middle, 1DC8), and BeF_3^--$NtrC^R$ (right, 1J56). Side chains of Asp54 (red), Thr82 (blue), and Tyr101 (magenta) are shown. The phosphate and BeF_3^- attached to Asp54 are shown in black. The fourth α-helix, which undergoes the most significant rearrangement when $NtrC^R$ is activated, has been labeled.

$NtrC^R$ undergoes a two state allosteric activation mechanism where the activation event, in this case phosphorylation, changes the relative energies of the active and inactive states rather than driving the protein into a completely new conformation (Nohaile et al. 1997). This results in a change in the populations of the two states, with a much higher percentage resembling the active conformation after phosphorylation. The chemical shift of specific residues for a set of mutants (V115I, D54E, D86N, D86N/A89T, D86N/A89T/V115I) all fall in a straight line between WT $NtrC^R$ and P-$NtrC^R$ and do so in order of increasing transcriptional activity (Volkman et al. 2001). This demonstrates that each of the mutants and the phosphorylated $NtrC^R$ itself are in fast exchange between an active and inactive state.

5.2.4 Positive and negative regulation of receiver domains

The receiver domain can act positively or negatively, depending on whether it promotes assembly of the oligomer once activated or inhibits assembly of the oligomer until activated (Chen et al. 2008). Phosphorylation of a positively regulated receiver domain

allows it to bind to and actively promote oligomerization of the hexamer, so the transcriptional activator can only be active in its presence. An unphopshorylated, negatively regulated receiver domain blocks the oligomerization of the active hexamer, so transcriptional activators are constitutively active in its absence.

For negatively regulated *A.a.* NtrC1 and NtrC4, phosphorylation of the receiver domain promotes a structural rearrangement that breaks the dimer interface between two adjacent receiver domains in the inactive dimer (Doucleff, Chen, et al. 2005)(Batchelor et al. 2008). This releases central domain subunits to reorganize, with two sets of dimers bound to adjacent enhancer sequences and one more dimer from solution oligomerizing into the active hexameric AAA+ ATPase. The absence of a receiver domain thus allows the active hexamer to form spontaneously in solution (Figure 5.3). These consitutively active transcriptional activators were used extensively in Chapter 2 to study the interaction between σ^{54} and the oligomeric transcriptional activators under the simplest conditions.

For positively regulated *E.c.* NtrC, the phosphorylated receiver domain undergoes a conformational rearrangement that exposes a hydrophobic surface (Kern et al. 1999). The surface binds to the central domain of the adjacent subunit and promotes oligomerization (De Carlo et al. 2006). Without the receiver domain, these positively regulated transcriptional activators cannot assemble (Figure 5.3). BeF_3^- or mutations to the receiver domain have been shown to mimic the phosphorylated state of the receiver domains *in vitro* (Klose, Weiss, and Kustu 1993)(Volkman et al. 1995)(Nohaile et al. 1997). In particular, when the natural phosphorylation site, aspartate 54, is substituted with glutamate the full length NtrC has constitutive activity *in vivo*.

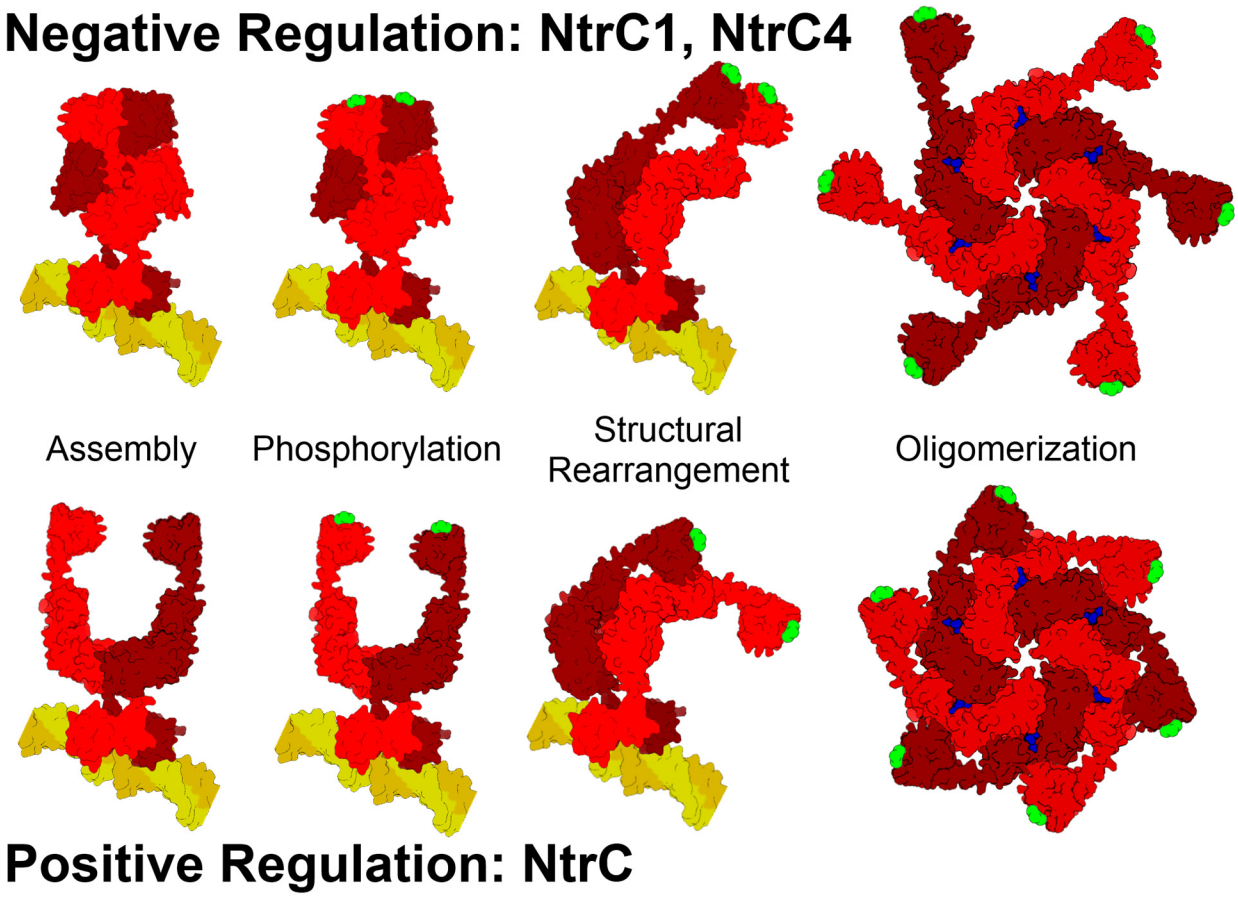


Figure 5.3 The steps of receiver domain activation with the NtrC homolog (red) and the enhancer DNA (yellow), phosphorylation (green) and ATP/ADP (blue). In both positive and negative regulation, activator dimers assemble at enhancer sites upstream of their gene target where the receiver domain is phosphorylated at an aspartate residue by the histidine kinase. This causes positively regulated receiver domains to bind the central domain of the adjacent subunit and promote oligomerization and negatively regulated receiver domains to alter their dimerization interface allowing oligomerization of the central ATPase domains. The receiver domains of negatively regulated transcriptional activators may remain in an alternative dimer configuration even after the central ATPase domains form hexamers, as in the case of $NtrC1^{R}$ (Doucleff, Chen, et al. 2005).

5.2.5 Pressure sensitive activation of *Shewanella violacea* $NtrC^{R}$

Like *E. coli*, the glnA gene from the piezophilic bacteria *Shewanella violacea* (Nogi, Kato, and Horikoshi 1998)(Kato and Nogi 2001) is regulated by two promoter sequences, one under control of σ^{70} (glnAp1) and the other under control of σ^{54} (glnAp2) (Reitzer and Magasanik 1985). The levels of the mRNA transcripts from glnAp1 are not affected by pressure, but the levels of mRNA transcripts from glnAp2 are much higher when cells are grown at 50 MPa instead of at atmospheric pressure, 0.1 MPa (Nakasone et al. 2002) (Figure 5.4).

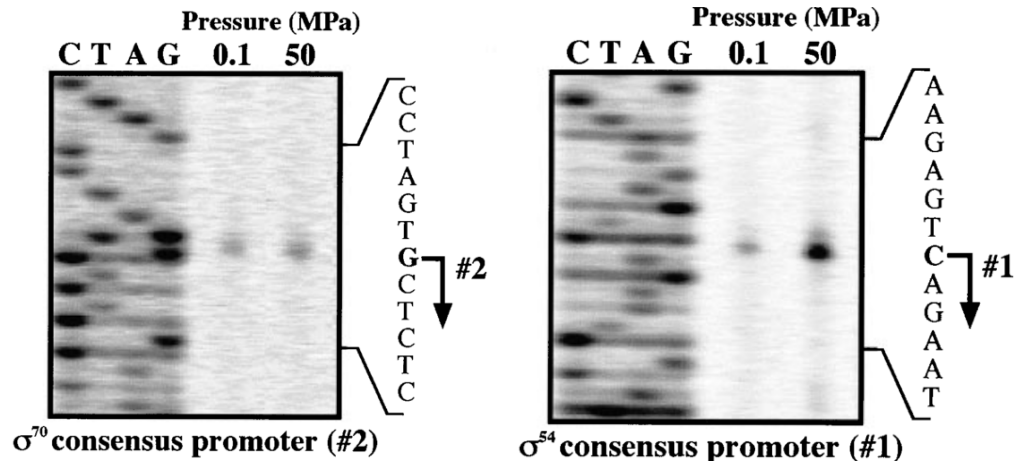


Figure 5.4 DNA gels showing the transcription product from the glnAp1 promoter (σ^{70}, left) and the glnAp2 promoter (σ^{54}, right) showing elevated levels of glnAp2 transcript at high pressure. Gels adapted from (Nakasone et al. 2002).

The concentration of NtrC is higher in cells cultured at 50 MPa relative to 0.1 MPa, while levels of σ^{54} are unchanged at different pressures. The glnALG operon which codes for glutamine synthetase (glnA), NtrB (glnL) and NtrC (glnG), is itself regulated by NtrC and σ^{54} as part of the two component signal transduction pathway (Sasse-Dwight and Gralla 1988). The increase in NtrC levels in cells cultured at high pressure indicates that one of the steps in the NtrC/σ^{54} two component signal transduction pathway in *S. violacea* is sensing pressure and initiating the cell's pressure response.

It is not clear which protein in the pathway is the pressure sensor that leads to increased transcription of NtrC-regulated genes. In *E. coli*, the NtrC receiver domain undergoes a large conformational rearrangement when phosphorylated that exposes a hydrophobic surface. This allosteric activation is brought about by a shift in the population of the active and inactive forms of the receiver domain (Volkman et al. 2001). Increased pressure has also been shown to lead to changes in the populations of protein folding intermediates (Kachel et al. 2006). We used high pressure NMR techniques in order to test if pressure itself is able to shift the receiver domain of NtrC into the active conformation without being phosphorylated.

5.2.6 Overview of high pressure NMR spectroscopy

High pressure NMR is an application of NMR spectroscopy that allows multidimensional NMR experiments to be undertaken at pressures up to 200 MPa. Pressure-resisting glass cells were developed in order to observe samples at high pressure use existing NMR magnets without drastically changing the entire probe head (Yamada et al. 2001)(Hiroaki Yamada 1974). The liquid samples are contained in thick-walled

NMR tubes made of borosilicate or quartz glass that is able to withstand 200 MPa of pressure but the inner diameter of the tube is only 1 mm leaving room for only 40 μL of sample. Sapphire cells have also been developed to withstand pressures up to 200 MPa with a wider inner diameter that is able to fit larger volumes of sample (Arnold, Kalbitzer, and Kremer 2003).

Raising the pressure produces two different responses in the protein. First, there is a structural change based on the local compression of the protein, primarily the shortening of amide hydrogen bonds, which results in amide proton chemical shift changes. Second, there is a shift in the equilibrium between different structural conformations (Li, Yamada, and Akasaka 1998). The second behavior has been applied to study sparsely populated intermediate states of proteins like human prion protein (PrP) (Kremer et al. 2007)(Kachel et al. 2006) and the relative stability of domains in Ras subjected to pressure induced unfolding (Inoue et al. 2000).

We studied the *S.v.* NtrC receiver domain to determine if changes in pressure alone could act as an environmental stimulus to signal a transcriptional responses to high pressure. The *S.v.* NtrC receiver domain exists in two states: an active conformation that is primarily occupied when it is phosphorylated and an inactive conformation when it is not. We examined the relative shift in these populations at high pressure to see if pressure itself could shift the equilibrium to the active conformation even without phosphorylation of the receiver domain. This type of regulation would bypass NtrB and immediately initiate transcription of pressure response genes known to be under control of *S.v.* NtrC and σ^{54}.

5.3 Results

5.3.1 *S.v.* $\mathbf{NtrC^R}$ sequence analysis

The sequence of *S. violacea* $NtrC^R$ much more closely resembles the *E. coli* $NtrC^R$ than *A. aeolicus* $NtrC1^R$ (Figure 5.5). *S. enterica* $NtrC^R$ used in some structural studies reported in this chapter is only three amino acids different (T31I, A37N, E41A) from *E.c.* $NtrC^R$ and assumed to be functionally identical. Like *E.c.* and *S.e.* NtrC, *S.v.* NtrC lacks the $TGXGXHydX_3HydX_2Hyd$ motif found in *A.a.* $NtrC1^R$ and *Sinorhizobium meliloti* DctD (Doucleff, Chen, et al. 2005). This conserved sequence is a mark of negatively regulated receiver domains and its absence in both *E.c.* and *S.v.* NtrC supports *S.v.* NtrC undergoing positive regulation.

```
A.a.        ---MNVLVIE DDKVFRGLLE EYLSMKGIKV ESAERGKEAY KLLSEKHFNV VLLDLLLPDV
S.v.        -MTEQVWVLD DDSSIRWVVE RALKGAKISS ASFAAAESLW EALELSQPKV IISDIRMPGT
E.c.        MQRGIVWVVD DDSSIRWVLE RALAGAGLTC TTFENGAEVL EALASKTPDV LLSDIRMPGM
S.e.        MQRGIVWVVD DDSSIRWVLE RALAGAGLTC ITFENGNEVL AALASKTPDV LLSDIRMPGM

A.a.        NGLEILKWIK ERSPETEVIV ITGHGTIKTA VEAMKMGAYD FLTKPCMLEE IELTINKAIE
S.v.        DGLSLLERLQ IHYPNIPVII MTAHSDLDSA VSAYQAGAFE YLPKPFDIDE AVTLVERALT
E.c.        DGLALLKQIK QRHPMLPVII MTAHSDLDAA VSAYQQGAFD YLPKPFDIDE AVALVERAIS
S.e.        DGLALLKQIK QRHPMLPVII MTAHSDLDAA VSAYQQGAFD YLPKPFDIDE AVALVERAIS

A.a.        HRK
S.v.        HST
E.c.        HYQ
S.e.        HYQ
```

Figure 5.5 Sequence alignment of NtrC receiver domain homologs from *Aquifex aeolicus*, *Shewanella violacea*, *Escherichia coli*, and *Salmonella enterica*. The activating phosphorylation site, D54, is shown in red. Dimerization motif of negatively regulated receiver domains, such as *A.a.* $NtrC^{R}$, is shown in bold.

Also the *S.v.* and *E.c.* NtrC receiver domain sequences are highly similar, containing 86% conserved amino acids and 65% identical amino acids. The conserved aspartate (D54) is the site of phosphorylation and suggests that *S.v.* NtrC is likely activated by a phosphotransfer from NtrB even if it also is able to bypass this activation at high pressure.

5.3.2 *S.v.* $NtrC^{R}$ undergoes structural rearrangements in the presence of BeF_3^-

^{15}N-labeled *S.v.* NtrC receiver domain was examined with ^{1}H-^{15}N HSQCs to examine its structural rearrangement upon activation. It produced a good spectrum with most peaks visible and little differential line broadening. The addition of BeF_3^-, which is able to mimic the phosphorylated state in *E.c.* $NtrC^{R}$ (Hastings et al. 2003), produced a significant change in chemical shifts (Figure 5.6).

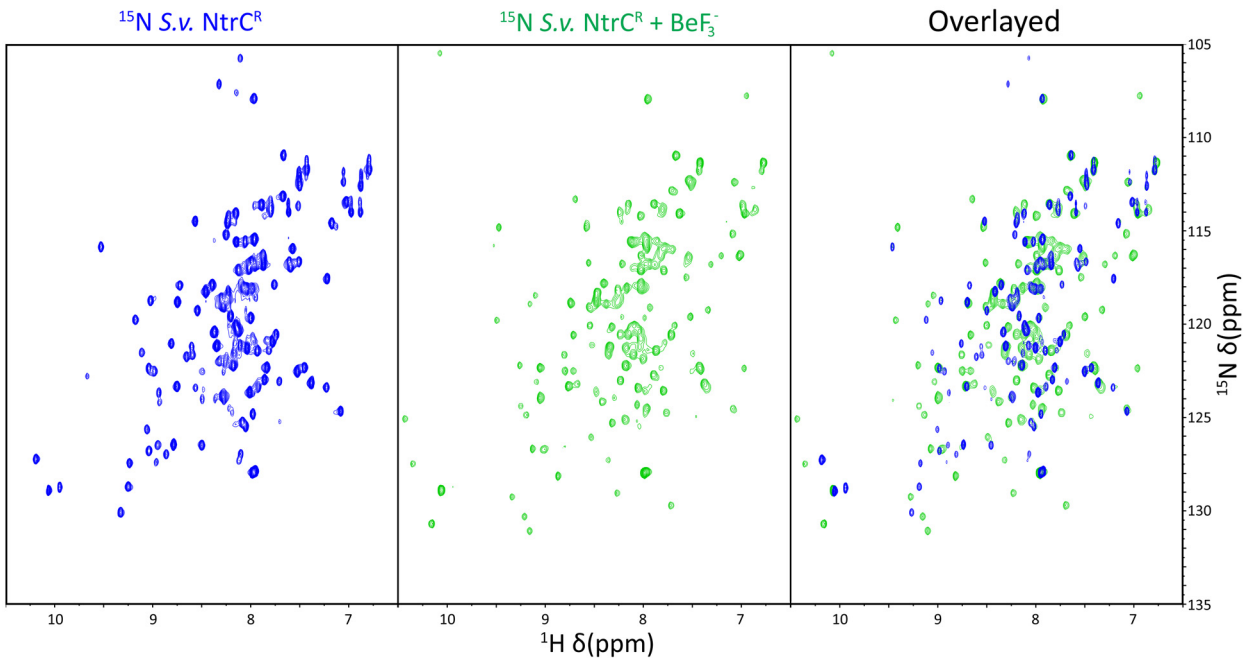


Figure 5.6 Activation of *S.v.* NtrC^{R} showing chemical shift changes of the amides in a ^{1}H-^{15}N HSQC spectrum in the absence of BeF_3^- (left, blue), the presence of a stoichiometric amount of BeF_3^- (center, green) and both overlayed (right).

Overlays of the ^{1}H-^{15}N HSQC spectra with and without BeF_3^- show a shift in the *S.v.* NtrC^{R} peaks. The number of peaks remains roughly the same indicating a shift in the population of the protein to the new conformation. Given the sequence similarity between *S.v.* and *E.c.* NtrC^{R} it is very likely that this newly populated conformation is the activated state.

5.3.3 Temperature induced chemical shift changes of *S.v.* NtrC^{R}

To study the effect of high and low temperatures on the energy landscape of *S.v.* NtrC^{R}, changes to the ^{1}H-^{15}N HSQC chemical shifts were observed at increasing temperatures (Figure 5.7). The receiver domain remained folded between the tested temperatures, 10°C to 30°C. At higher temperatures, amide chemical shifts moved downfield in the hydrogen dimension but showed only small changes in the nitrogen dimension. The amide peak trajectories followed a linear path with temperature, indicating a shift in the populations of various conformations of the protein in a fast exchanging equilibrium.

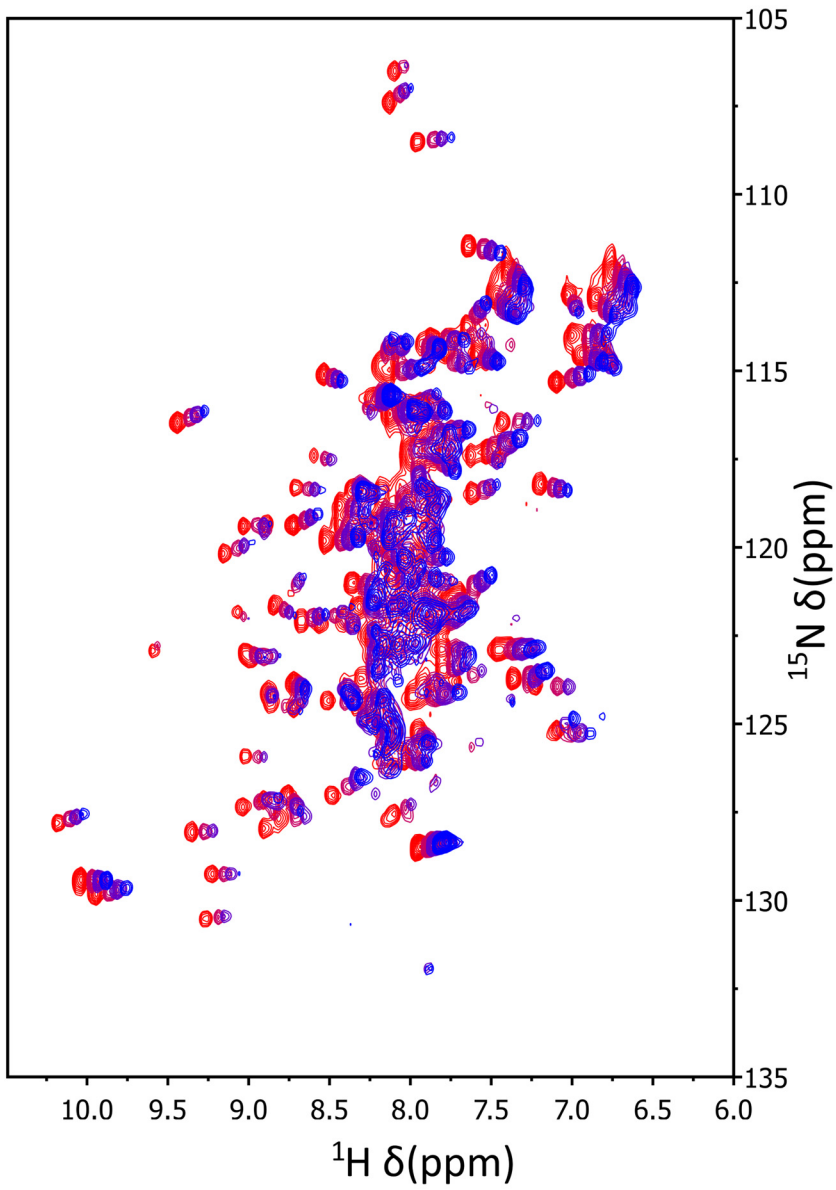


Figure 5.7 Temperature induced chemical shift changes of *S.v.* NtrC^{R} from a ^{1}H-^{15}N HSQC. 30°C (red), 20°C (red-violet), 15°C (blue-violet), 10°C (blue). Peak movement is linear with temperature.

5.3.4 Pressure induced conformational changes of *S.v* NtrC^{R}

To determine if increasing pressure produces the same state as BeF_3^- activation, ^{1}H-^{15}N HSQCs of ^{15}N-labeled *S.v.* NtrC^{R} were taken at increasing pressures in 10 MPa increments up to 200 MPa with all other experimental conditions kept the same (Figure 5.8).

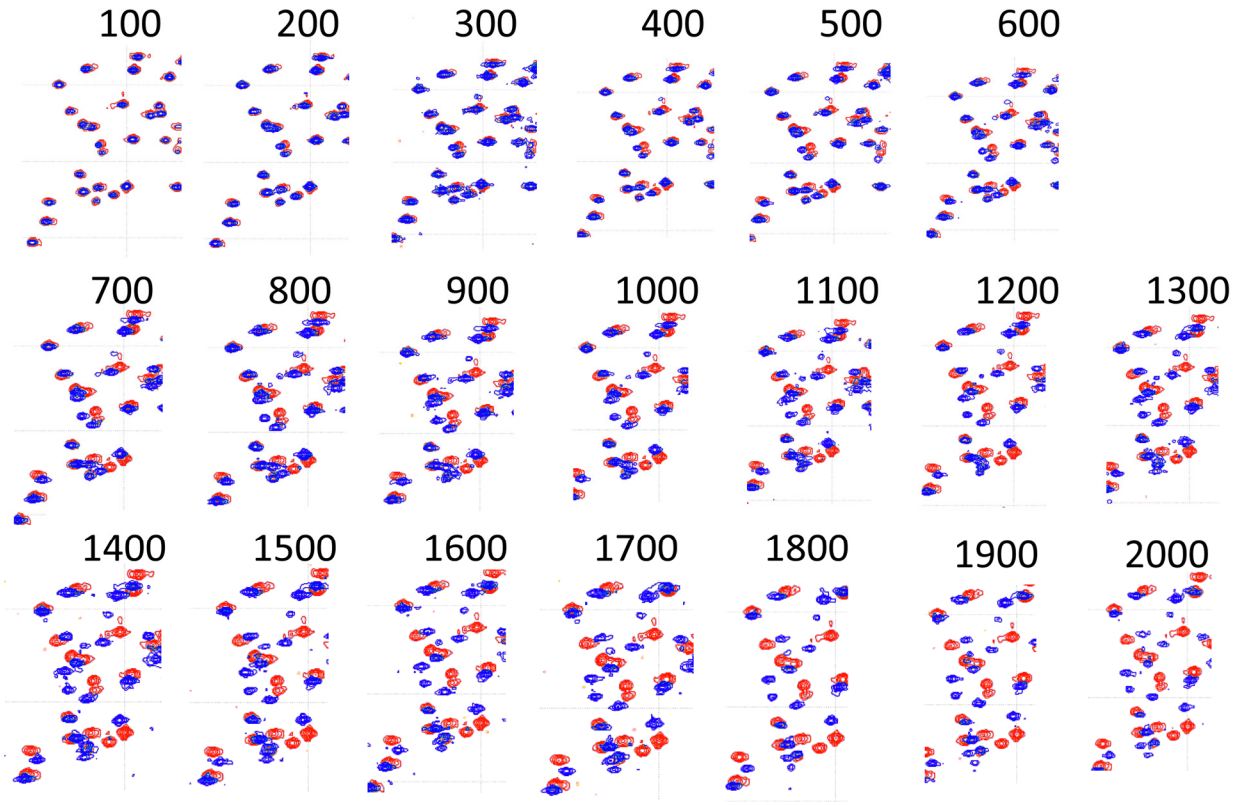
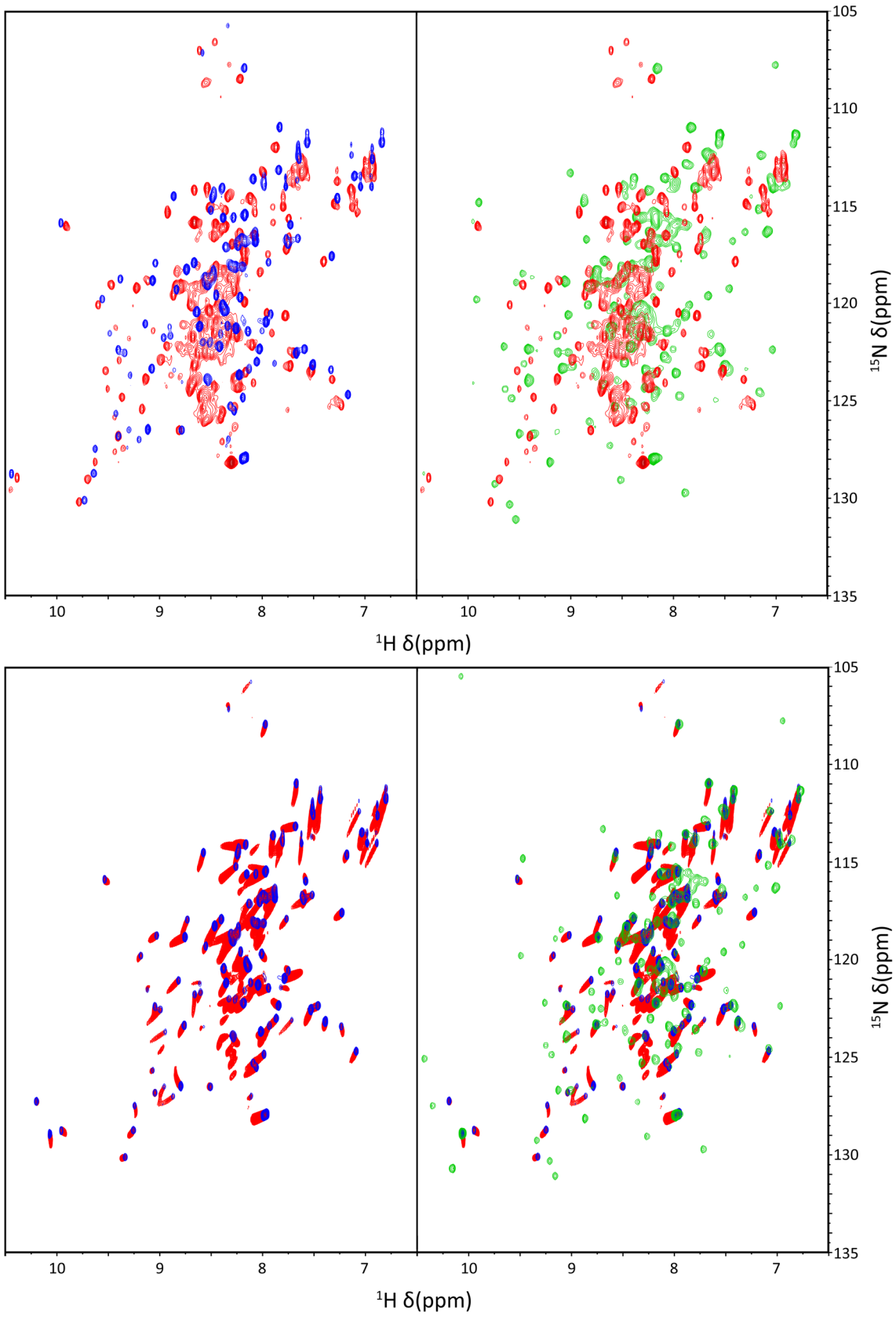


Figure 5.8 Chemical shift changes of the ^{1}H-^{15}N HSQC spectra of *S.v.* NtrC^{R} at increase pressures from 100 bar (10 MPa) to 2000 bar (200 MPa) in red, overlayed with the spectrum of *S.v.* NtrC^{R} at atmospheric pressure (0.1 MPa) in blue. High pressure NMR spectra collected by Werner Kremer.

Unlike pressure induced unfolding of other proteins like Ras at similar pressures (Inoue et al. 2000), *S.v.* NtrC^{R} peaks shift but new peaks do not appear in a manner that would suggest significant unfolding of the protein even locally (Figure 5.9). Though there is some degree of differential peak broadening, very few peaks are broadened beyond detection even at 200 MPa. The chemical shift dispersion indicates that the protein remains folded at high pressure.

The trajectory of the chemical shift changes as pressure increases suggests a simple two state equilibrium in fast exchange. As pressure increases, the vast majority of chemical shifts lie on a collinear path towards higher ppm in both the ^{1}H and ^{15}N dimensions (Figure 5.9).


105
110
115
120
125
130
135
^{15}N δ(ppm)
10
9
8
7
^{1}H δ(ppm)

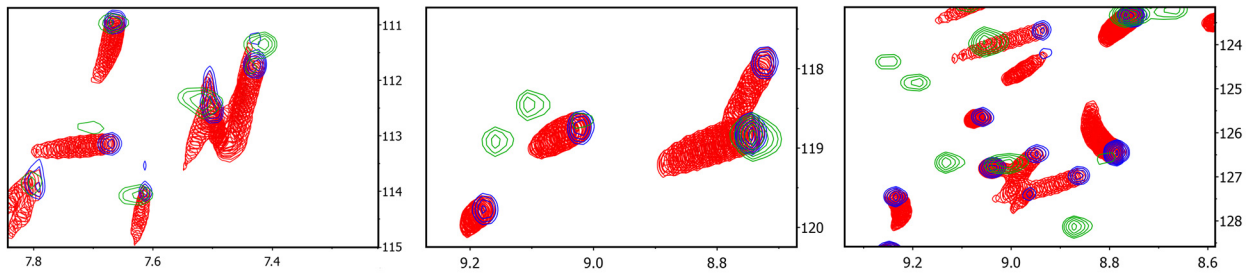


Figure 5.9 Top: Overlayed ^{1}H-^{15}N HSQCs of ^{15}N *S.v.* NtrC^{R} taken at atmospheric pressure (blue), atmospheric pressure with BeF_3^- activation (green), and 200 MPa pressure (red). Middle: The trajectory of the peak shifts as pressure increases with an overlay of all pressures in 10 MPa increments from 10 to 200 MPa (red) overlayed with the spectrum at atmospheric pressure (blue) and atmospheric pressure with BeF_3^- activation (green). Bottom: Zoomed in regions of the spectrum show that increasing pressure causes linear chemical shift changes with a few exceptions (right most spectrum) but does not shift peaks toward the chemical shifts in the BeF_3^- activated *S.v.* NtrC^{R} spectrum (green).

However, the chemical shifts of the high pressure conformation do not match the activated conformation of *S.v.* NtrC^{R} found at excess concentrations of the phosphate mimic BeF_3^-. Similarly, the pressure induced chemical shift change trajectories do not match those brought about by raising or lowering the temperature (Figure 5.7).

5.4 Discussion

S.v. NtrC is implicated in the regulation of σ^{54} transcription of genes involved in pressure response. Levels of *S.v.* NtrC itself are elevated in cells grown at pressures up to 50 MPa or 500 times atmospheric pressure (Nakasone et al. 2002)(Nogi, Kato, and Horikoshi 1998). This indicates that a component of the NtrB/NtrC two component signal transduction pathway is involved in sensing pressure and driving the accumulation of activated NtrC that in turn initiates transcription of σ^{54} and the genes that control the cell's pressure response. In this chapter, we examined whether the regulatory receiver domain of NtrC could be the pressure sensing protein, bypassing its activation by NtrB which is used to initiate other cellular responses like those needed to survive on limited nitrogen.

5.4.1 S.v. NtrC Receiver Domain is likely positively regulated

The receiver domain of *S.v.* NtrC has high sequence homology with positively regulated *Salmonella enterica* NtrC, including a nearly identical α-helix four, which undergoes the most significant rearrangement when the receiver domain is driven into the active conformation. *S.v.* NtrC lacks the TGXGXHydX$_3$HydX$_2$Hyd motif found in *A.a.* NtrC1

that is a determinant of negative regulation (Doucleff, Chen, et al. 2005). It's therefore likely that *S.v.* NtrC is positively regulated and that upon adopting the active state conformation it binds to the central domain of NtrC and promotes the assembly of the active NtrC hexamer.

S.e. $NtrC^{R}$ has been shown to undergo allosteric activation rather than an induced fit (Volkman et al. 2001). This means that instead of the phosphorylation event inducing a conformational change in the receiver domain, it changes the energy landscape and shifts the equilibrium between two states: from an overwhelming majority of molecules folded in the inactive conformation to a majority folded in the active conformation. High pressure NMR has been used to study sparsely populated intermediate states by changing their population (Kachel et al. 2006)(Kremer et al. 2007). It therefore seemed possible that high pressure could increase the population of *S.v.* $NtrC^{R}$ in the active state even without being phosphorylated. To test this hypothesis, we observed the NMR chemical shift changes of unactivated *S.v.* $NtrC^{R}$ in ^{1}H-^{15}N HSQCs at increasing pressures as a measure of its structural similarity to the active conformation.

5.4.2 High pressure conformational changes of the *S.v.* NtrC receiver domain

Amide peaks in ^{1}H-^{15}N HSQCs were seen to shift along a linear trajectory as the pressure is increased in 10 MPa increments from 0.1 MPa up to 200 MPa. Most but not all of the chemical shift changes follow a straight line suggesting that *S.v.* $NtrC^{R}$ conformations are in fast exchange and experiencing a steadily changing energy landscape, and thus changing equilibrium populations, as the pressure increases.

The chemical shifts of the amide protons move downfield in both the ^{1}H and ^{15}N dimensions as pressure increases, which is consistent with existing knowledge of local changes to proteins under pressure (Akasaka, Tezuka, and Yamada 1997). These changes are primarily due to a compression of the solvent and shortening of the hydrogen bonding between amides and their binding partners: carbonyls on nearby amino acids or water for solvent-exposed amides (Li, Yamada, and Akasaka 1998).

Higher temperatures and pressures can both drive denaturation and both cause the amide proton chemical shifts to move downfield (Heremans and Smeller 1998). This is consistent with the trajectory of most but not all *S.v.* $NtrC^{R}$ peaks in both the pressure and temperature experiments, which could be interpreted as an increasing population of

unfolded proteins. Unlike temperature denaturation, which is generally a cooperative two-state unfolding where the protein exists in equilibrium between an unfolded and folded state, high pressure denaturation often shows significant deviations from two-state unfolding. High pressure denaturation is attributed to compression of internal cavities in the folded state that result in gradual breaking of native contacts (Roche et al. 2012). At high enough pressures, this eventually leads to unfolding in a highly variable manner that depends on each protein's unique energy landscape. The linear trajectory of most of the chemical shift changes of *S.v.* $NtrC^{R}$ under increasing pressure is more consistent with a gradual conformational change than high pressure unfolding. The most likely source of the conformational change is a shortening of hydrogen bond lengths in the folded protein at high pressure, which would lead to a gradual but consistent change in the chemical environment of the amides participating in hydrogen bonding and would account for the observed linear trajectory of amide chemical shift changes. The small shift changes and remaining chemical shift dispersion suggest that the piezophilic protein has remained folded up to 200 MPa, consistent with a protein from an organism that has evolved to survive at high pressures.

5.4.3 *S.v.* NtrC receiver domain does not refold into the active conformation at high pressure

To determine if the two-state behavior observed in *S.v.* $NtrC^{R}$ is related to an increasingly populated active conformation of the receiver domain at high pressures, we compared the pressure induced chemical shift changes to the chemical shifts of the *S.v.* $NtrC^{R}$ phosphorylated state mimic using the phosphate analog BeF_3^-. Titrating in BeF_3^- to *S.v.* $NtrC^{R}$ at atmospheric pressure resulted in significant chemical shift changes to the amide ^{1}H-^{15}N HSQC. These changes are similar in magnitude to the changes in *S.e.* $NtrC^{R}$ when it is phosphorylated or bound to BeF_3^-, which indicate that the BeF_3^--activated *S.v.* $NtrC^{R}$ is very likely the active conformation of the receiver domain (Kern et al. 1999)(Hastings et al. 2003). However, the chemical shift changes brought on by increasing the pressure do not correlate well with those brought on by activating with BeF_3^-. Even the direction the amide peaks move at increasing pressure does not correlate with the BeF_3^--activated state. This suggests that high pressure is not shifting the equilibrium from the inactive state to the active state. It is likely that the primary contributor to the observed chemical shift changes at high pressure is the shrinking hydrogen bond distances, rather than an unfolding event or a conformational change towards the activate state of *S.v.* $NtrC^{R}$ and thus these observations do not support pressure induced activation. The *S. violacea* pressure response in transcription is most

likely derived from a protein other than NtrC, with the most obvious candidates being NtrB or P_{II} though a third protein unique to piezophiles cannot be ruled out (Inoue et al. 2000).

5.4.4 Future Directions

It is unlikely that the NtrC receiver domain itself is the pressure sensing mechanism that drives transcription of pressure response genes in *Shewanella violacea*. It is possible that the high pressure conformation, while not itself active, is more readily phosphorylated or that its phosphorylated state is less quickly autodephosphorylated. These would both lead to a buildup of the active, phosphorylated state of NtrC which would in turn initiate transcription of pressure response genes in the standard fashion. However, it is also possible that NtrB or P_{II} sense the pressure and lead to the phosphorylation and activation of the NtrC receiver domain.

To determine the pressure sensing mechanism of *S. violacea*, we would need to make constructs of *S.v.* NtrB and possibly also P_{II} and observe their phosphotransfer and dephosphorylation activity *in vitro* at high pressures. This would be challenging given that *E.c.* NtrC-P autodephosphorylates on its own in 5 minutes, which is too quick for most NMR experiments (Keener and Kustu 1988). Quenching the reactions by chemically denaturing NtrC with urea immediately after incubation with NtrB and ATP at high and low pressures may be necessary to observe the phosphotransfer to NtrC and isolate this system's pressure sensor.

5.5 Materials and Methods

5.5.1 *S.v.* $\mathbf{NtrC^R}$ plasmid sequence

The *Shewanella violacea* $NtrC^R$(1-123) DNA sequence was synthesized as an IDT-DNA gBlocks fragment to match the known protein sequence. However, the DNA sequence itself was chosen using the most common *E. coli* codons, rather than the actual sequence of the *S.v.* $NtrC^R$ gene:

```
CAGTCATATGACCGAACAGGTTTGGGTTCTGGATGATGATAGCAGCATTCGTTGGGTTGTTGAACGTGCACTGAAAGGTGCAAAAATTAGCAGCGCA
AGCTTTGCAGCAGCAGAAAGCCTGTGGGAAGCACTGGAACTGAGCCAGCCGAAAGTGATTATTAGCGATATTCGTATGCCTGGCACCGATGGTCTGA
GCCTGCTGGAACGTCTGCAGATTCATTATCCGAATATTCCGGTTATTATCATGACCGCACATAGCGATCTGGATAGCGCAGTTAGCGCATATCAGGC
AGGCGCATTTGAATATCTGCCGAAACCGTTTGATATTGATGAAGCAGTTACCCTGGTGGAACGTGCCCTGACCCATAGCACCGATTAAGGATCCATC
G
```

The fragment was amplified with PCR primers matching the 5' and 3' ends of the above sequence. PCR product was cut with NdeI and BamHI and ligated into pET-29a vectors behind a His_6 tag cleavable by TEV protease.

5.5.2 Protein preparation

Plasmids were transformed into BL21(DE3) Rosetta cells, grown at 37°C (in LB or ^{15}N M9 medium) and induced at an OD_{600nm} of 0.6 with 1 mM IPTG and grown at 37°C for 3 more hours. Cells were sonicated and pelleted. *S.v.* $NtrC^R$ protein went to inclusion bodies and remained in the pellet. Pelleted protein was recovered by resuspending the pellet in Column Buffer (8M Urea, 10 mM Tris, 100 mM sodium phosphate) at pH 8.0 and pelleting a second time, this time taking the supernatant and loading onto an NiNTA column.

S.v. $NtrC^R$ was purified on an NiNTA column under denaturing conditions by washing with Column Buffer at pH 8.0 followed by Column Buffer at pH 6.3 and eluting with Column Buffer at pH 4.5. The elution was diluted to 40 mL with Column Buffer at pH 8.0 and dialyzed overnight into NMR Buffer (50 mM sodium phosphate, 50 mM NaCl, pH 6.75, 1 mM DTT). TEV protease was added for one additional night, and cut *S.v.* $NtrC^R$ was passed through an NiNTA column run at native conditions, coming out in the flow through.

5.5.3 Activation by beryllium fluoride

S.v. $NtrC^R$ protein was activated by the addition of beryllium fluoride. A beryllium fluoride solution was made by adding 20-40 mM $BeCl_2$ to NMR Buffer and mixing in 8 times the molar concentration of NaF until solution becomes clear. Then a slight excess (25-45 mM) of a stoichiometric amount of $MgCl_2$ was added and the precipitate was filtered by passing through a 0.2 μm filter membrane. Stock solutions of BeF_3^- were made with between 20-40 mM of beryllium and added directly to the NMR tube to achieve final concentrations in stoichiometric excess of the protein concentration. Though other beryllium fluoride compounds like BeF_4^{2-} are present, BeF_3^- is the phosphate mimic found in crystal structures and is used to represent beryllium fluoride throughout this work.

5.5.4 NMR Spectroscopy at atmospheric and high pressures

S.v. $NtrC^{R}$ was concentrated to 400 μM in NMR Buffer + 10% D_2O and observed on a Bruker Avance 800 MHz spectrometer at 303K and 1 atm (0.1 MPa). Chemical shift changes due to activation were observed with a ^{1}H-^{15}N HSQC. Data were processed with NMRPipe (Delaglio et al. 1995) and chemical shift analysis was undertaken with CARA (Keller 2004) or MestreNOVA (Cobas and Sardina 2003).

High pressure HSQC experiments were undertaken in collaboration with Werner Kremer and Hans-Robert Kalbitzer at Universitat Regensburg, Germany. ^{15}N-labeled protein was lyophilized and hydrated back into the original buffer with H_2O + 10% D_2O. 1H-15N HSQCs were taken on a Bruker Avance 800 MHz spectrometer at 303K and between 0.1 and 200 MPa. Data was processed with NMRPipe and chemical shift analysis was undertaken with CARA or MestreNOVA.

References

Akasaka, Kazuyuki, Tomoko Tezuka, and Hirokaki Yamada. 1997. "Pressure-Induced Changes in the Folded Structure of Lysozyme." *Journal of Molecular Biology* 271 (5): 671–78.

Arnold, Martin Reinhard, Hans Robert Kalbitzer, and Werner Kremer. 2003. "High-Sensitivity Sapphire Cells for High Pressure NMR Spectroscopy on Proteins." *Journal of Magnetic Resonance* 161 (2): 127–31.

Baba, Tomoya, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. 2006. "Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection." *Molecular Systems Biology* 2 (1).

Barne, Kerry A., Jon A. Bown, Stephen J. W. Busby, and Stephen D. Minchin. 1997. "Region 2.5 of the Escherichia Coli RNA Polymerase σ70 Subunit Is Responsible for the Recognition of the 'Extended -10' Motif at Promoters." *The EMBO Journal* 16 (13): 4034–40.

Barrios, Humberto, Brenda Valderrama, and Enrique Morett. 1999. "Compilation and Analysis of σ-Dependent Promoter Sequences." *Nucleic Acids Research* 27 (22): 4305–13.

Batchelor, Joseph D., Michaeleen Doucleff, Chul-Jin Lee, Koshi Matsubara, Sacha De Carlo, Johanna Heideker, Meindert H Lamers, Jeffrey G Pelton, and David E Wemmer. 2008. "Structure and Regulatory Mechanism of Aquifex Aeolicus NtrC4: Variability and Evolution in Bacterial Transcriptional Regulation." *Journal of Molecular Biology* 384 (5): 1058–75.

Batchelor, Joseph D., Peter S. Lee, Andrew C. Wang, Michaeleen Doucleff, and David E. Wemmer. 2013. "Structural Mechanism of GAF-Regulated σ54 Activators from Aquifex Aeolicus." *Journal of Molecular Biology* 425 (1): 156–70.

Batchelor, Joseph D., Harry J Sterling, Eunmi Hong, Evan R Williams, and David E Wemmer. 2009. "Receiver Domains Control the Active-State Stoichiometry of Aquifex Aeolicus σ54 Activator NtrC4, as Revealed by Electrospray Ionization Mass Spectrometry." *Journal of Molecular Biology* 393 (3): 634–43.

Beier, Dagmar, and Roy Gross. 2006. "Regulation of Bacterial Virulence by Two-Component Systems." *Current Opinion in Microbiology* 9 (2): 143–52.

Bose, Daniel, Tillmann Pape, Patricia C Burrows, Mathieu Rappas, Siva R Wigneshweraraj, Martin Buck, and Xiaodong Zhang. 2008. "Organization of an Activator-Bound RNA Polymerase Holoenzyme." *Molecular Cell* 32 (3): 337–46.

Bourret, Robert B, Katherine a Borkovich, and Melvin I Simon. 1991. "Pathways

Involving Protein Phosphorylation in Prokaryotes." *Annual Review of Biochemistry* 60: 401–41.

Brahmachary, Priyanka, Mona G. Dashti, Jonathan W. Olson, and Timothy R. Hoover. 2004. "Helicobacter Pylori FlgR Is an Enhancer-Independent Activator of σ54-RNA Polymerase Holoenzyme." *Journal of Bacteriology* 186 (14): 4535–42.

Browning, Douglas F, and Stephen J W Busby. 2004. "The Regulation of Bacterial Transcription Initiation." *Nature Reviews Microbiology* 2 (1): 57–65.

Buchan, Daniel W A, Federico Minneci, Tim C O Nugent, Kevin Bryson, and David T. Jones. 2013. "Scalable Web Services for the PSIPRED Protein Analysis Workbench." *Nucleic Acids Research* 41 (Web server issue): 349–57.

Buck, Martin, Daniel Bose, Patricia C Burrows, Wendy Cannon, Nicolas Joly, Tillmann Pape, Mathieu Rappas, Jörg Schumacher, Siva R Wigneshweraraj, and Xiaodong Zhang. 2006. "A Second Paradigm for Gene Activation in Bacteria." *Biochemical Society Transactions* 34 (6): 1067–71.

Buck, Martin, Maria-Trinidad Gallegos, David J Studholme, Yuli Guo, and Jay D Gralla. 2000. "MINIREVIEW The Bacterial Enhancer-Dependent σ54 (σN) Transcription Factor." *Journal of Bacteriology* 182 (15): 4129–36.

Burgess, Richard R, Andrew A Travers, John J Dunn, and Ekkehard K F Bautz. 1969. "Factor Stimulating Transcription by RNA Polymerase." *Nature* 221 (5175): 43–46.

Bustamante, Carlos, Yann R. Chemla, Nancy R. Forde, and David Izhaky. 2004. "Mechanical Processes in Biochemistry." *Annual Review of Biochemistry* 73 (1): 705–48.

Campbell, Elizabeth A, Lars F Westblade, and Seth A Darst. 2008. "Regulation of Bacterial RNA Polymerase σ Factor Activity: A Structural Perspective." *Current Opinion in Microbiology* 11 (2): 121–27.

Campbell, Elizabeth A., Oriana Muzzin, Mark Chlenov, Jing L. Sun, C. Anders Olson, Oren Weinman, Michelle L. Trester-Zedlitz, and Seth A. Darst. 2002. "Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit." *Molecular Cell* 9 (3): 527–39.

Cecconi, Ciro, Elizabeth A. Shank, Carlos Bustamante, and Susan Marqusee. 2005. "Direct Observation of the Three-State Folding of a Single Protein Molecule." *Science* 309 (5743): 2057–60.

Cecconi, Ciro, Elizabeth A. Shank, Frederick W. Dahlquist, Susan Marqusee, and Carlos Bustamante. 2008. "Protein-DNA Chimeras for Single Molecule Mechanical Folding Studies with the Optical Tweezers." *European Biophysics Journal* 37 (6): 729–38.

Chaney, Matthew K, Ricardo Grande, Siva R Wigneshweraraj, Wendy Cannon, Paul Casaz, Maria-Trinidad Gallegos, Jörg Schumacher, et al. 2001. "Binding of

Transcriptional Activators to Sigma 54 in the Presence of the Transition State Analog ADP-Aluminum Fluoride: Insights into Activator Mechanochemical Action." *Genes & Development* 15 (17): 2282–94.

Chen, Baoyu, Michaeleen Doucleff, David E Wemmer, Sacha De Carlo, Hector H Huang, E Nogales, Timothy R Hoover, E Kondrashkina, Liang Guo, and B Tracy Nixon. 2007. "ATP Ground- and Transition States of Bacterial Enhancer Binding AAA plus ATPases Support Complex Formation with Their Target Protein, σ54." *Structure* 15 (4): 429–40.

Chen, Baoyu, Tatyana A Sysoeva, Saikat Chowdhury, Liang Guo, Sacha De Carlo, Jeffrey A. Hanson, Haw Yang, and B Tracy Nixon. 2010. "Engagement of Arginine Finger to ATP Triggers Large Conformational Changes in NtrC1 AAA+ ATPase for Remodeling Bacterial RNA Polymerase." *Structure (London, England : 1993)* 18 (11): 1420–30.

Chen, Baoyu, Tatyana A. Sysoeva, Saikat Chowdhury, and B. Tracy Nixon. 2008. "Regulation and Action of the Bacterial Enhancer-Binding Protein AAA+ Domains." *Biochemical Society Transactions* 36 (Pt 1): 89–93.

Cherepanov, Peter P., and Wilfried Wackernagel. 1995. "Gene Disruption in Escherichia Coli: TcR and KmR Cassettes with the Option of Flp-Catalyzed Excision of the Antibiotic-Resistance Determinant." *Gene* 158 (1): 9–14.

Claverie-Martin, F, and B Magasanik. 1991. "Role of Integration Host Factor in the Regulation of the glnHp2 Promoter of Escherichia Coli." *Proc. Natl. Acad. Sci. USA* 88 (5): 1631–35.

Cobas, Juan Carlos, and F. Javier Sardina. 2003. "Nuclear Magnetic Resonance Data Processing. MestRe-C: A Software Package for Desktop Computers." *Concepts in Magnetic Resonance* 19A (2): 80–96.

Colland, Frédéric, Gilbert Orsini, Edward N. Brody, Henri Buc, and Annie Kolb. 1998. "The Bacteriophage T4 AsiA Protein: A Molecular Switch for Sigma 70-Dependent Promoters." *Molecular Microbiology* 27 (4): 819–29.

Dago, Angel Ernesto, Siva R Wigneshweraraj, Martin Buck, and Enrique Morett. 2007. "A Role for the Conserved GAFTGA Motif of AAA plus Transcription Activators in Sensing Promoter DNA Conformation." *Journal of Biological Chemistry* 282 (2): 1087–97.

Datsenko, K A, and B L Wanner. 2000. "One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products." *Proc. Natl. Acad. Sci. USA* 97 (12): 6640–45.

De Carlo, Sacha, Baoyu Chen, Timothy R. Hoover, Elena Kondrashkina, Eva Nogales, and B. Tracy Nixon. 2006. "The Structural Basis for Regulated Assembly and Function of the Transcriptional Activator NtrC." *Genes & Development* 20 (11): 1–

7.

Delaglio, Frank, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and Ad Bax. 1995. "NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes." *Journal of Biomolecular NMR* 6 (3): 277–93.

Delagoutte, Emmanuelle, and Peter H von Hippel. 2003. "Helicase Mechanisms and the Coupling of Helicases within Macromolecular Machines - Part II: Integration of Helicases into Cellular Processes." *Quarterly Reviews of Biophysics* 36 (1): 1–69.

Dietz, Hendrik, and Matthias Rief. 2004. "Exploring the Energy Landscape of GFP by Single-Molecule Mechanical Experiments." *Proc. Natl. Acad. Sci. USA* 101 (46): 16192–97.

Dill, Jesse. 2012. "Struts, Springs and Crumple Zones: Protein Structures under Force."

Doucleff, Michaeleen, Baoyu Chen, Ann E Maris, David E Wemmer, Elena Kondrashkina, and B Tracy Nixon. 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." *Journal of Molecular Biology* 353 (2): 242–55.

Doucleff, Michaeleen, Lawrence T Malak, Jeffrey G Pelton, and David E Wemmer. 2005. "The C-Terminal RpoN Domain of σ54 Forms an Unpredicted Helix-Turn-Helix Motif Similar to Domains of σ70." *Journal of Biological Chemistry* 280 (50): 41530–36.

Doucleff, Michaeleen, Jeffrey G Pelton, Peter S Lee, B Tracy Nixon, and David E Wemmer. 2007. "Structural Basis of DNA Recognition by the Alternative Sigma-Factor, σ54." *Journal of Molecular Biology* 369 (4): 1070–78.

Dunker, A. Keith, Celeste J. Brown, J. David Lawson, Lilia M. Iakoucheva, and Zoran Obradovic. 2002. "Intrinsic Disorder and Protein Function." *Proteins* 41 (21): 6573–82.

Dyson, H. Jane, and Peter E Wright. 2004. "Unfolded Proteins and Protein Folding Studied by NMR." *Chemical Reviews* 104 (8): 3607–22.

Gallegos, Maria-Trinidad, and Martin Buck. 2000. "Sequences in σ54 Region I Required for Binding to Early Melted DNA and Their Involvement in Sigma-DNA Isomerisation." *Journal of Molecular Biology* 297 (4): 849–59.

Ghosh, Tamaswati, Daniel Bose, and Xiaodong Zhang. 2010. "Mechanisms for Activating Bacterial RNA Polymerase." *FEMS Microbiology Reviews* 34 (5): 611–27.

Glynn, Steven E, Andreas Martin, Andrew R Nager, Tania a Baker, and Robert T Sauer. 2009. "Structures of Asymmetric ClpX Hexamers Reveal Nucleotide-Dependent Motions in a AAA+ Protein-Unfolding Machine." *Cell* 139 (4): 744–56.

Gnatt, Averell L, Patrick Cramer, Jianhua Fu, David A. Bushnell, and Roger D.

Kornberg. 2001. “Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution.” *Science* 292 (5523): 1876–83.

Goto, Natalie K., Kevin H. Gardner, Geoffrey A. Mueller, Randall C. Willis, and Lewis E. Kay. 1999. “A Robust and Cost-Effective Method for the Production of Val, Leu, Ile (δ1) Methyl-Protonated 15N-, 13C-, 2H-Labeled Proteins.” *Journal of Biomolecular NMR* 13 (4): 369–74.

Greenleaf, William J., Michael T. Woodside, Elio A. Abbondanzieri, and Steven M. Block. 2005. “Passive All-Optical Force Clamp for High-Resolution Laser Trapping.” *Physical Review Letters* 95 (20): 1–4.

Gyaneshwar, Prasad, Oleg Paliy, Jon McAuliffe, David L Popham, Michael I Jordan, and Sydney Kustu. 2005. “Sulfur and Nitrogen Limitation in Escherichia Coli K-12: Specific Homeostatic Responses.” *J. Bacteriol.* 187 (3): 1074–90.

Hastings, Curtis a, Seok-Yong Lee, Ho S Cho, Dalai Yan, Sydney Kustu, and David E Wemmer. 2003. “High-Resolution Solution Structure of the Beryllofluoride-Activated NtrC Receiver Domain.” *Biochemistry* 42 (30): 9081–90.

Heremans, K, and L Smeller. 1998. “Protein Structure and Dynamics at High Pressure.” *Biochimica et Biophysica Acta* 1386 (2): 353–70.

Hong, Eunmi, Michaeleen Doucleff, and David E Wemmer. 2009. “Structure of the RNA Polymerase Core-Binding Domain of σ54 Reveals a Likely Conformational Fracture Point.” *Journal of Molecular Biology* 390 (1): 70–82.

Howard, J., A. J. Hudspeth, and R. D. Vale. 1989. “Movement of Microtubules by Single Kinesin Molecules.” *Nature* 342 (6246): 154–58.

Hsieh, Mingli, and Jay D Gralla. 1994. “Analysis of the N-Terminal Leucine Heptad and Hexad Repeats of Sigma 54.” *Journal of Molecular Biology* 239 (1): 15–24.

Hsieh, Mingli, Yin Tintut, and Jay D Gralla. 1994. “Functional Roles for the Glutamines within the Glutamine-Rich Region of the Transcription Factor σ54.” *The Journal of Biological Chemistry* 269 (1): 373–78.

Hummer, Gerhard, and Attila Szabo. 2001. “Free Energy Reconstruction from Nonequilibrium Single-Molecule Pulling Experiments.” *Proc. Natl. Acad. Sci. U. S. A.* 98 (7): 3658.

Inoue, Kyoko, Hiroaki Yamada, Kazuyuki Akasaka, Christian Herrmann, Werner Kremer, Till Maurer, Rolf Döker, and Hans Robert Kalbitzer. 2000. “Pressure-Induced Local Unfolding of the Ras Binding Domain of RalGDS.” *Nature Structural Biology* 7 (7): 547–50.

Jones, David T. 1999. “Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices.” *J. Mol. Biol.* 292: 195–202.

Juang, Yue-Li, and John D Helmann. 1994. “A Promoter Melting Region in the Primary σ Factor of Bacillus Subtilis. Identification of Functionally Important Aromatic

Amino Acids." *Journal of Molecular Biology* 235 (5): 1470–88.

Kachel, Norman, Werner Kremer, Ralph Zahn, and Hans Robert Kalbitzer. 2006. "Observation of Intermediate States of the Human Prion Protein by High Pressure NMR Spectroscopy." *BMC Structural Biology* 6 (16).

Kato, Chiaki, and Yuichi Nogi. 2001. "Correlation between Phylogenetic Structure and Function: Examples from Deep-Sea Shewanella." *FEMS Microbiology Ecology* 35 (3): 223–30.

Kazmierczak, Mark J, Martin Wiedmann, and Kathryn J Boor. 2005. "Alternative Sigma Factors and Their Roles in Bacterial Virulence." *Microbiology and Molecular Biology Reviews* 69 (4): 527–43.

Keener, John, and Sydney Kustu. 1988. "Protein Kinase and Phosphoprotein Phosphatase Activities of Nitrogen Regulatory Proteins NTRB and NTRC of Enteric Bacteria: Roles of the Conserved Amino-Terminal Domain of NTRC." *Proc. Natl. Acad. Sci. USA* 85 (14): 4976–80.

Keller, R. 2004. *The Computer Aided Resonance Assignment Tutorial. Goldau, Switzerland.*

Kern, Dorothee, Brian F Volkman, Peter Luginbühl, Michael J Nohaile, Sydney Kustu, and David E Wemmer. 1999. "Structure of a Transiently Phosphorylated Switch in Bacterial Signal Transduction." *Nature* 402 (6764): 894–98.

Klose, Karl E, David S Weiss, and Sydney Kustu. 1993. "Glutamate at the Site of Phosphorylation of Nitrogen-Regulatory Protein NTRC Mimics Aspartyl-Phosphate and Activates the Protein." *Journal of Molecular Biology* 232 (1): 67–78.

Korzheva, Nataliya, Arkady Mustaev, Maxim Kozlov, Arun Malhotra, Vadim Nikiforov, and Seth A. Darst Alex Goldfarb. 2000. "A Structural Model of Transcription Elongation." *Science* 289 (5479): 619–25.

Kremer, Werner, Norman Kachel, Kazuo Kuwata, Kazuyuki Akasaka, and Hans Robert Kalbitzer. 2007. "Species-Specific Differences in the Intermediate States of Human and Syrian Hamster Prion Protein Detected by High Pressure NMR Spectroscopy." *The Journal of Biological Chemistry* 282 (31): 22689–98.

Kustu, Sydney, Eduardo Santero, John Keener, David Popham, and David Weiss. 1989. "Expression of σ54 (ntrA)-Dependent Genes Is Probably United by a Common Mechanism." *Microbiological Reviews* 53 (3): 367–76.

Kyte, Jack, and Russell F. Doolittle. 1982. "A Simple Method for Displaying the Hydropathic Character of a Protein." *Journal of Molecular Biology* 157 (1): 105–32.

Lee, Seok-Yong, Armando De La Torre, Dalai Yan, Sydney Kustu, B Tracy Nixon, and David E Wemmer. 2003. "Regulation of the Transcriptional Activator NtrC1: Structural Studies of the Regulatory and AAA+ ATPase Domains." *Genes & Development* 17 (20): 2552–63.

Li, Hua, Hiroaki Yamada, and Kazuyuki Akasaka. 1998. "Effect of Pressure on Individual Hydrogen Bonds in Proteins. Basic Pancreatic Trypsin Inhibitor." *Biochemistry* 37 (5): 1167–73.

Lisser, Shlomit, and Hanah Margalit. 1993. "Compilation of E. Coli mRNA Promoter Sequences." *Nucleic Acids Research* 21 (7): 1507–16.

Liu, Cuihua, and Craig T. Martin. 2002. "Promoter Clearance by T7 RNA Polymerase. Initial Bubble Collapse and Transcript Dissociation Monitored by Base Analog Fluorescence." *Journal of Biological Chemistry* 277 (4): 2725–31.

Liu, Jiangang, Narayanan B Perumal, Christopher J Oldfield, Eric W Su, Vladimir N Uversky, and A Keith Dunker. 2006. "Intrinsic Disorder in Transcription Factors," 6873–88.

Lonetto, Michael, Michael Gribskov, and Carol A Gross. 1992. "The σ70 Family: Sequence Conservation and Evolutionary Relationships." *Journal of Bacteriology* 174 (12): 3843–49.

Losick, Richard, and Janice Pero. 1981. "Cascades of Sigma Factors." *Cell* 25 (September): 582–84.

Malhotra, Arun, Elena Severinova, and Seth A Darst. 1996. "Crystal Structure of a σ70 Subunit Fragment from E. Coli RNA Polymerase." *Cell* 87 (1): 127–36.

Maniatis, Tom, Edward F. Fritsch, and Joseph Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*.

Matouschek, Andreas. 2003. "Protein Unfolding - An Important Process in Vivo?" *Current Opinion in Structural Biology* 13 (1): 98–109.

Mizuno, T. 1997. "Compilation of All Genes Encoding Two-Component Phosphotransfer Signal Transducers in the Genome of Escherichia Coli." *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes* 4 (2): 161–68.

Murakami, Katsuhiko S, Katsuhiko S Murakami, Shoko Masuda, and Seth A Darst. 2002. "Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution." *Science* 296 (5571): 1280–84.

Murakami, Katsuhiko S., and Seth A. Darst. 2003. "Bacterial RNA Polymerases: The Wholo Story." *Current Opinion in Structural Biology* 13 (1): 31–39.

Nakasone, Kaoru, Akihiko Ikegami, Hiroaki Kawano, Chiaki Kato, Ron Usami, and Koki Horikoshi. 2002. "Transcriptional Regulation under Pressure Conditions by RNA Polymerase σ54 Factor with a Two-Component Regulatory System in Shewanella Violacea." *Extremophiles* 6 (2): 89–95.

Neuwald, Andrew F., L. Aravind, John L. Spouge, and Eugene V. Koonin. 1999. "AAA+: A Class of Chaperone-like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes." *Genome Research* 9 (1): 27–43.

Ninfa, Alexander J., and Richard L. Bennett. 1991. "Identification of the Site of Autophosphorylation of the Bacterial Protein Kinase/phosphatase NRII." *Journal of Biological Chemistry* 266 (11): 6888–93.

Ninfa, Alexander J., Lawrence J. Reitzer, and Boris Magasanik. 1987. "Initiation of Transcription at the Bacterial glnAp2 Promoter by Purified E. Coli Components Is Facilitated by Enhancers." *Cell* 50 (7): 1039–46.

Nixon, B Tracy, Clive W Ronson, and Frederick M Ausubel. 1986. "Two-Component Regulatory Systems Responsive to Environmental Stimuli Share Strongly Conserved Domains with the Nitrogen Assimilation Regulatory Genes ntrB and ntrC." *Proc. Natl. Acad. Sci. USA* 83 (20): 7850–54.

Nogi, Yuichi, Chiaki Kato, and Koki Horikoshi. 1998. "Taxonomic Studies of Deep-Sea Barophilic Shewanella Strains and Description of Shewanella Violacea Sp. Nov." *Archives of Microbiology* 170 (5): 331–38.

Nohaile, M, D Kern, D Wemmer, K Stedman, and S Kustu. 1997. "Structural and Functional Analyses of Activating Amino Acid Substitutions in the Receiver Domain of NtrC: Evidence for an Activating Surface." *Journal of Molecular Biology* 273 (1): 299–316.

Nohno, Tsutomu, and Taiichi Saito. 1987. "Two Transcriptional Start Sites Found in the Promoter Region of Escherichia Coli Glutamine Permease Operon, glnHPQ." *Nucleic Acids Res* 15 (6): 2777.

North, a K, and S Kustu. 1997. "Mutant Forms of the Enhancer-Binding Protein NtrC Can Activate Transcription from Solution." *Journal of Molecular Biology* 267 (1): 17–36.

Oguiza, Jose A, Maria-Trinidad Gallegos, Matthew K Chaney, Wendy Cannon, and Martin Buck. 1999. "Involvement of the σN DNA-Binding Domain in Open Complex Formation." *Molecular Microbiology* 33 (4): 873–85.

Paget, Mark, and John Helmann. 2003. "Protein Family Review: The σ70 Family of Sigma Factors." *Genome Biology* 4 (1): 1–6.

Park, Sungdae, Matthew Meyer, a Daniel Jones, Hemant P Yennawar, Neela H Yennawar, and B Tracy Nixon. 2002. "Two-Component Signaling in the AAA + ATPase DctD: Binding Mg2+ and BeF3- Selects between Alternate Dimeric States of the Receiver Domain." *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 16 (14): 1964–66.

Pelton, J G, S Kustu, and D E Wemmer. 1999. "Solution Structure of the DNA-Binding Domain of NtrC with Three Alanine Substitutions." *Journal of Molecular Biology* 292 (5): 1095–1110.

Pitt, M, Maria-Trinidad Gallegos, and Martin Buck. 2000. "Single Amino Acid Substitution Mutants of Klebsiella Pneumoniae sigma(54) Defective in

Transcription." *Nucleic Acids Research* 28 (22): 4419–27.

Porter, Susan C., Anne K. North, Andrew B. Wedel, and Sydney Kustu. 1993. "Oligomerization of NTRC at the glnA Enhancer Is Required for Transcriptional Activation." *Genes and Development* 7 (11): 2258–73.

Rappas, Mathieu, Daniel Bose, and Xiaodong Zhang. 2007. "Bacterial Enhancer-Binding Proteins: Unlocking σ54-Dependent Gene Transcription." *Current Opinion in Structural Biology* 17 (1): 110–16.

Rappas, Mathieu, Jorg Jörg Schumacher, Fabienne Beuron, Hajime Niwa, Patricia Bordes, Sivaramesh R Wigneshweraraj, Catherine A Keetch, Carol V Robinson, Martin Buck, and Xiaodong Zhang. 2005. "Structural Insights into the Activity of Enhancer-Binding Proteins." *Science* 307 (5717): 1972–75.

Rappas, Mathieu, Jörg Schumacher, Hajime Niwa, Martin Buck, and Xiaodong Zhang. 2006. "Structural Basis of the Nucleotide Driven Conformational Changes in the AAA+ Domain of Transcription Activator PspF." *J. Mol. Biol.* 357: 481–92.

Reitzer, Lawrence J. 2003. "Nitrogen Assimilation and Global Regulation in Escherichia Coli." *Annual Review of Microbiology* 57: 155–76.

Reitzer, Lawrence J, and Boris Magasanik. 1985. "Expression of glnA in Escherichia Coli Is Regulated at Tandem Promoters." *Proc. Natl. Acad. Sci. USA* 82 (7): 1979–83.

Roche, Julien, Jose A Caro, Douglas R Norberto, Philippe Barthe, Christian Roumestand, Jamie L Schlessman, Angel E Garcia, Bertrand E García-Moreno, and Catherine A Royer. 2012. "Cavities Determine the Pressure Unfolding of Proteins." *Proc Natl Acad Sci USA* 109 (18): 6945–50.

Ronson, Clive W, B. Tracy Nixon, and Frederick M. Ausubel. 1988. "Conserved Domains in Bacterial Regulatory Proteins That Respond to Environmental Stimuli." *Cell* 49: 579–81.

Ross, Wilma, Khoosheh K. Gosink, Julia Salomon, Kazuhiko Igarashi, Chao Zou, Akira Ishihama, Konstantin Severinov, and Richard L. Gourse. 1993. "Third Recognition Element in Bacterial Promoters: DNA Binding by the Subunit of RNA Polymerase." *Science* 262 (5138): 1407–13.

Sallai, László, and Paul A. Tucker. 2005. "Crystal Structure of the Central and C-Terminal Domain of the σ54-Activator ZraR." *Journal of Structural Biology* 151 (2): 160–70.

Santero, Eduardo, Timothy R. Hoover, Anne K. North, David K. Berger, Susan C. Porter, and Sydney Kustu. 1992. "Role of Integration Host Factor in Stimulating Transcription from the sigma54-Dependent nifH Promoter." *Journal of Molecular Biology* 227 (3): 602–20.

Sasse-Dwight, Selina, and Jay D Gralla. 1988. "Probing the Escherichia Coli glnALG Upstream Activation Mechanism in Vivo." *Proc. Natl Acad. Sci. USA* 85 (23):

8934–38.

Schulz, Alexandra, Jörg Langowski, and Karsten Rippe. 2000. "The Effect of the DNA Conformation on the Rate of NtrC Activated Transcription of Escherichia Coli RNA Polymerase-σ54 Holoenzyme." *Journal of Molecular Biology* 300 (4): 709–25.

Shank, Elizabeth a, Ciro Cecconi, Jesse W Dill, Susan Marqusee, and Carlos Bustamante. 2010. "The Folding Cooperativity of a Protein Is Controlled by Its Chain Topology." *Nature* 465 (7298): 637–40.

Sheng, Wanyun, and Xiubei Liao. 2002. "Solution Structure of a Yeast Ubiquitin-like Protein Smt3: The Role of Structurally Less Defined Sequences in Protein-Protein Recognitions." *Protein Science* 11 (6): 1482–91.

Skordalakes, Emmanuel, and James M. Berger. 2006. "Structural Insights into RNA-Dependent Ring Closure and ATPase Activation by the Rho Termination Factor." *Cell* 127 (3): 553–64.

Smith, S., Cui, Y., Bustamante, C. 2003. "Optical-Trap Force Transducer That Operates by Direct Measurement of Light Momentum." *Methods Enzymol.* 361: 134–62.

Southern, E, and Mike J. Merrick. 2000. "The Role of Region II in the RNA Polymerase σ Factor σN (σ54)." *Nucleic Acids Research* 28 (13): 2563–70.

Studholme, David J, and Ray Dixon. 2003. "Domain Architectures of σ54-Dependent Transcriptional Activators." *Journal of Bacteriology* 185 (6): 1757–67.

Studier, F. William. 2014. "Stable Expression Clones and Auto-Induction for Protein Production in E. Coli." In *Structural Genomics*, edited by Yu Wai Chen, 1091:17–32.

Syed, Adeela, and Jay D Gralla. 1998. "Identification of an N-Terminal Region of Sigma 54 Required for Enhancer Responsiveness." *J. Bacteriol* 180 (21): 5619–25.

Sysoeva, Tatyana A., Saikat Chowdhury, Liang Guo, B Tracy Nixon, B. Tracy Nixon, and B Tracy Nixon. 2013. "Nucleotide-Induced Asymmetry within ATPase Activator Ring Drives σ54-RNAP Interaction and ATP Hydrolysis." *Genes & Development* 27 (22): 2500–2511.

Tahirov, Tahir H, Dmitry Temiakov, Michael Anikin, Vsevolod Patlan, William T McAllister, Dmitry G Vassylyev, and Shigeyuki Yokoyama. 2002. "Structure of a T7 RNA Polymerase Elongation Complex at 2.9Å Resolution." *Nature* 420: 43–50.

Tantos, Agnes, Kyou-Hoon Han, and Peter Tompa. 2012. "Intrinsic Disorder in Cell Signaling and Gene Transcription." *Molecular and Cellular Endocrinology* 348 (2): 457–65.

Thomsen, Nathan D., and James M. Berger. 2009. "Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases." *Cell* 139 (3): 523–34.

Tucker, Nicholas P, Tamaswati Ghosh, Matthew Bush, Xiaodong Zhang, and Ray

Dixon. 2009. “Essential Roles of Three Enhancer Sites in σ54-Dependent Transcription by the Nitric Oxide Sensing Regulatory Protein NorR.” *Nucleic Acids Research* 38 (4): 1182–94.

Tucker, Paul A., and László Sallai. 2007. “The AAA+ Superfamily - a Myriad of Motions.” *Current Opinion in Structural Biology* 17 (6): 641–52.

Tugarinov, Vitali, Peter M Hwang, Jason E Ollerenshaw, and Lewis E Kay. 2003. “Cross-Correlated Relaxation Enhanced 1H-13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes.” *Journal of the American Chemical Society* 125 (34): 10420–28.

Tyler, Bonnie. 1978. “Regulation of the Assimilation of the Nitrogen Compounds.” *Annual Review of Biochemistry* 47: 1127–62.

Ulrich, Luke E., Eugene V. Koonin, and Igor B. Zhulin. 2005. “One-Component Systems Dominate Signal Transduction in Prokaryotes.” *Trends in Microbiology* 13 (2): 52–56.

Vallee, Richard B., John C. Williams, Dileep Varma, and Lora E. Barnhart. 2004. “Dynein: An Ancient Motor Protein Involved in Multiple Modes of Transport.” *Journal of Neurobiology* 58 (2): 189–200.

Vassylyev, Dmitry G, Shun-ichi Sekine, Oleg Laptenko, Jookyung Lee, Marina N Vassylyeva, Sergei Borukhov, and Shigeyuki Yokoyama. 2002. “Crystal Structure of a Bacterial RNA Polymerase Holoenzyme at 2.6 Å Resolution.” *Nature* 417 (6890): 712–19.

Vidangos, Natasha, Ann E Maris, Anisa Young, Eunmi Hong, Jeffrey G Pelton, Joseph D Batchelor, and David E Wemmer. 2013. “Structure, Function, and Tethering of DNA-Binding Domains in σ(54) Transcriptional Activators.” *Biopolymers* 99 (12): 1082–96.

Volkman, Brian F, D Lipson, D E Wemmer, and D Kern. 2001. “Two-State Allosteric Behavior in a Single-Domain Signaling Protein.” *Science* 291 (5512): 2429–33.

Volkman, Brian F, Michael J Nohaile, J- Nancy K Amy, Sydney Kustu, David E Wemmer, N K Amy, Sydney Kustu, and David E Wemmer. 1995. “Three-Dimensional Solution Structure of the N-Terminal Receiver Domain of NTRC.” *Biochemistry* 34 (4): 1413–24.

Wang, Jonathan T, Adeela Syed, and Jay D Gralla. 1997. “Multiple Pathways to Bypass the Enhancer Requirement of Sigma 54 RNA Polymerase: Roles for DNA and Protein Determinants.” *Proc. Natl. Acad. Sci. USA* 94 (18): 9538–43.

Wang, Jonathan T, Adeela Syed, Mingli Hsieh, and Jay D Gralla. 1995. “Converting Escherichia Coli RNA Polymerase into an Enhancer-Responsive Enzyme: Role of an NH2-Terminal Leucine Patch in σ54.” *Science* 270 (5238): 992–94.

Wang, Michelle D, Mark J Schnitzer, H Yin, Robert Landick, Jeff Gelles, and Steven M

Block. 1998. “Force and Velocity Measured for Single Molecules of RNA Polymerase.” *Science* 282 (5390): 902–7.

Wedel, Andrew B, and Sydney Kustu. 1995. “The Bacterial Enhancer-Binding Protein NTRC Is a Molecular Machine: ATP Hydrolysis Is Coupled to Transcriptional Activation.” *Genes & Development* 9 (16): 2042–52.

Wedel, Andrew B, David S. Weiss, David Popham, Peter Droge, and Sydney Kustu. 1990a. “A Bacterial Enhancer Functions to Tether a Transcriptional Activator Near a Promoter.” *Science* 248 (4954): 486–90.

Wedel, Andrew B., David S. Weiss, David L Popham, Peter Droge, and Sydney Kustu. 1990b. “A Bacterial Enhancer Functions to Tether a Transcriptional Activator near a Promoter.” *Science* 248 (4954): 486–90.

Weiss, Verena, and Boris Magasanik. 1988. “Phosphorylation of Nitrogen Regulator I (NRI) of Escherichia Coli.” *Proc. Natl. Acad. Sci. USA* 85 (23): 8919–23.

West, Ann H., and Ann M. Stock. 2001. “Histidine Kinases and Response Regulator Proteins in Two-Component Signaling Systems.” *Trends in Biochemical Sciences* 26 (6): 369–76.

Wigneshweraraj, Siva R., Sergei Nechaev, Patricia Bordes, Susan Jones, Wendy Cannon, Konstantin Severinov, and Martin Buck. 2003. “Enhancer-Dependent Transcription by Bacterial RNA Polymerase: The β Subunit Downstream Lobe Is Used by σ54 during Open Promoter Complex Formation.” *Methods in Enzymology* 370 (2003): 646–57.

Wishart, David S, Brian D Sykes, and Frederic M Richards. 1991. “Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure.” *Journal of Molecular Biology* 222 (2): 311–33.

Wong, Cai’ne, Jay D Gralla, and Yin Tintut. 1994. “The Domain Structure of Sigma 54 as Determined by Analysis of a Set of Deletion Mutants.” *Journal of Molecular Biology* 236 (1): 81–90.

Wright, Peter E, and H Jane Dyson. 2015. “Intrinsically Disordered Proteins in Cellular Signalling and Regulation.” *Nature Publishing Group* 16 (1): 18–29.

Wuite, Gijs J.L., Steven B. Smith, Mark Young, David Keller, and Carlos Bustamante. 2000. “Single-Molecule Studies of the Effect of Template Tension on T7 DNA Polymerase Activity.” *Nature* 404 (6773): 103–6.

Wüthrich, Kurt. 1986. *NMR of Proteins and Nucleic Acids*. Edited by John Wiley & Sons.

Xiao, Yan, Siva R Wigneshweraraj, Robert Weinzierl, Yi-Ping Wang, and Martin Buck. 2009. “Construction and Functional Analyses of a Comprehensive σ54 Site-Directed Mutant Library Using Alanine-Cysteine Mutagenesis.” *Nucleic Acids Research* 37 (13): 4482–97.

Yamada, H., K. Nishikawa, M. Honda, T. Shimura, K. Akasaka, and K. Tabayashi. 2001. "Pressure-Resisting Cell for High-Pressure, High-Resolution Nuclear Magnetic Resonance Measurements at Very High Magnetic Fields." *Review of Scientific Instruments* 72 (2): 1463–71.

Yamada, Hiroaki. 1974. "Pressure-Resisting Glass Cell for High Pressure, High Resolution NMR Measurement" 45: 640–42.

Yang, Yun, Vidya C Darbari, Nan Zhang, Duo Lu, Robert Glyde, Yi-ping Wang, Jared T Winkelman, et al. 2015. "Structures of the RNA Polymerase-σ54 Reveal New and Conserved Regulatory Strategies." *Science* 349 (6250): 882–86.

Zhang, X., Matthew K Chaney, Siva R. Wigneshweraraj, Jörg Schumacher, P. Bordes, Wendy Cannon, and Martin Buck. 2002. "Mechanochemical ATPases and Transcriptional Activation." *Molecular Microbiology* 45 (4): 895–903.

Zuo, Yuhong, and Thomas A. Steitz. 2015. "Crystal Structures of the E. Coli Transcription Initiation Complexes with a Complete Bubble." *Molecular Cell* 58 (3): 534–40.