Katie Amberg-Johnson
Honors Thesis

Use of Molecular Tweezers to Study the Effect of Applied Force on σ^{54} Core-Binding Domain

Abstract

Transcription initiation by bacterial σ^{54}-RNA polymerase holoenzyme requires activation by an AAA+ ATPase activator protein. An interaction with the N-terminal region of sigma54 shifts the RNAP holoenzyme into an open complex where it can open its target DNA. This activation mechanism may involve a tugging force by the activator with resulting conformational changes to one or both subdomains of the core-binding domain causing a rearrangement to the holoenzyme. To investigate this, we use molecular tweezers to study the folding cooperativity of the σ^{54} core-binding domain under applied force. Our analysis reveals two unfolding transitions, the first accounts for 20% of the domain and the second, for the remaining 80%. Based on the size of the first transition, there is evidence that the C-terminal helix unfolds separately under applied force. We did not notice any unfolding transitions that might suggest non-cooperative unfolding of the two core-binding subdomains.

Introduction

In bacteria, transcription initiation requires binding of the core RNA polymerase (RNAP) to the promoter sequence, as well as the binding of the σ factor, a protein responsible for melting the double helix DNA to create room for RNA synthesis. Many different σ factors have been identified and each of them target a subset of genes for which they can initiate transcription. This step thus affords specificity and regulation to transcription [1].

The major class of σ factors, called σ^{70} in *Escherichia coli,* is responsible for initiating transcription of "housekeeping" genes responsible for normal cell growth. Upon binding the promoter sequence, σ^{70}-RNAP holoenzyme can immediately melt the double-stranded DNA and start synthesizing RNA [2]. The other class of σ factors, known as σ^{54}, is responsible for regulation of a variety of genes associated with nitrogen metabolism and virulence determinants in some animal pathogens. For example, both *Borrelia burgdorferi*, the agent of Lyme disease, and *Burkholderia cenocepacia*, the agent of respiratory infections in cystic fibrosis patients, require σ^{54} for survival in the human host [3,4]. σ^{54} shares no sequence homology with σ^{70} and uses a different mechanism [5].

σ^{54}-RNAP holoenzyme binds to the promoter DNA, but unlike other sigma factors, cannot open it without an additional activation step. An interaction between the N-terminal domain of σ^{54} and one of the AAA+ ATPase transcriptional activator proteins, followed by one or more rounds of ATP hydrolysis, is required to melt the DNA and initiate transcription [6]. This extra activation requirement makes σ^{54}-regulated gene expression more tightly controlled than σ^{70}-regulated genes.

σ^{54} is made up of four functional domains (Fig. 1). Starting from the N-terminus is the Activator Binding Domain (ABD), which interacts with the AAA+ ATPase activator proteins. Next is the Core-Binding Domain (CBD) which interacts with the subunits of core RNAP. The last two domains are DNA Binding Domains (DBD) that interact with the -12 DNA region and the -24 DNA region respectively. The -12DBD is involved in DNA melting while the -24DBD functions in positioning σ^{54} at the appropriate region along the DNA [7].

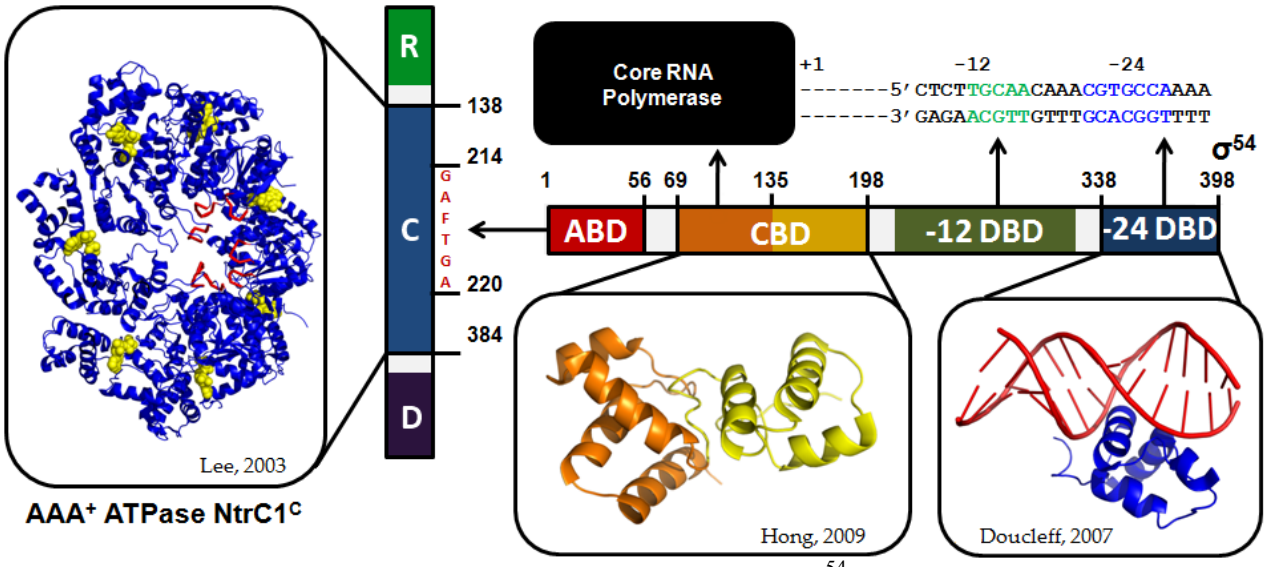

Fig. 1. This schematic shows the domains of Aquifex aeolicus σ^{54}, the elements (activator proteins, RNAP, or DNA) that interact with σ^{54}, and select solved structures. The yellow and orange sections of the CBD highlight the two subdomains.

Aquifex aeolicus CBD spans residues 69-198 and contains a four α-helix bundle (residues 69-135) as well as a three α-helix bundle (residues 135-198). The four helix bundle can fold on its own, and makes most of the contacts with core RNAP [8]. The three helix bundle packs tightly against the four helix bundle and cannot fold without it. The interface between the two domains is a conserved structural feature and could function as a fracture point that holds the two domains together until the activator protein exerts a large enough "tugging" force on the N-terminus to separate them [8]. This resulting conformational change could be the switch that triggers open complex formation and transcription initiation.

Materials and Methods

DNA Handle Generation

pGMEX-1 plasmid DNA was transformed into E. Coli Top 10 cells. One colony was grown overnight in Luria Broth (LB) media and 20 μg of plasmid were isolated using QIAGEN Prep Spin Miniprep Kit. 400 μg of the 558 base pair DNA handles were generated by PCR using 16.2 mM DTT, 1x PCR buffer, 1.5 mM magnesium chloride, 0.2 mM dNTPs, 0.5 uM of each primer, 0.04 ng of plasmid, 0.025 U Taq polymerase and 8.064 mL of water for a total of 10 mL PCR reaction. Reactions were transferred into a GeneMate 96-well 0.2 mL PCR plate flat and covered with an AlumaSeal 2 sheet and then placed in a 96-well thermocycler. Both handles were generated with the primer 5' thiol-GCT-ACC-GTA-ATT-GAG-ACC-AC and either 5' biotin-

CAA-AAA-ACC-CCT-CAA-GAC-CC or 5' digoxigenin-CAA-AAA-ACC-CCT-CAA-GAC-CC. The PCR products were purified using the QIAGEN Hi-Speed Plasmid Midi Kit.

Protein Preparation

DNA encoding a construct of the core-binding domain of A. aeolicus σ^{54}, residues 69-198 plus a $(His)_6$-tag at the N terminus, was transformed into E. coli Rosetta cells. Cells were grown in one liter of LB and induced at A_{600nm}=0.6 with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Three hours later, cells were collected by centrifugation and lysed in 20 mM sodium phosphate, 0.5 M NaCl, and 2 mM β-mercaptoethanol at pH 7.4 by sonication. Cell lysates were centrifuged at 30,000g for 30 minutes. The supernatant was heated to 70°C for 30 minutes and then spun overnight at 4°C on a nickel column. 50 mg of protein was purified from the Ni column using 20 mM sodium phosphate, 0.5 M NaCl, 2 mM β-mercaptoethanol and 20 mM imidazole at pH 7.4 to wash the column and 20 mM sodium phosphate, 0.5 M NaCl, 2 mM β-mercaptoethanol and 300 mM imidazole at pH 7.4 in the elution buffer.

DTDP Activation of Protein

Purified CBD was denatured by diluting 1000x in 6 M guandinium hydrochloride, and 0.1 M sodium phosphate at pH 7 and reconcentrating in 10kDa MWCO Millipore tubes. Proteins were reduced with 10 mM dithiothreitol (DTT) for 1 hour at room temperature in a 25 molar excess of dithiodipyridine (DTDP). Next they were buffer-exchanged into a buffer containing another 25 molar excess of DTDP and 0.1 M sodium phosphate at pH 5.5 by gravity filtration using Sephadex G-25 columns and left to react overnight at room temperature.

Attachment of DNA Handles to Protein

160 nmoles of each DNA handle (1:1) were concentrated to 100 μL total volume and reduced in a 40:1 molar excess of DTT for 1 hour at room temperature. They were then buffer exchanged using Bio-Rad Biospin 6 Chromatography columns into 0.1 M sodium phosphate buffer at pH 8. 80 nmoles of DTDP-activated protein in 0.05% tween was immediately added to the DNA handles and allowed to react over night at room temperature. Successful DNA handle-protein reactions were verified by native gel electrophoresis.

Mechanical Manipulation of the DNA-modified Proteins using the Molecular Tweezers

1 μL of resulting DNA-protein constructs were reacted with a solution of polystyrene beads coated in anti-digoxigenin antibodies (anti-dig) for 10 minutes at room temperature. This solution was flowed into the molecular tweezer trap under a microscope. One anti-dig bead at a time was brought in close proximity with another polystyrene bead coated in streptavidin held in place by a pipette until a tether between the two beads was attained. Moving the laser and optically trapped bead relative to the bead held by the pipette results in stretching, unfolding, or refolding of the protein. Five single protein molecules were tested at six different unfolding forces and over 20 force-extension cycles each.

Results

This construct of CBD contains alanines in place of cysteines 94 and 119 as well as cysteines in place of the first and last amino acids in the sequence. Cysteines form the disulfide bond between

the DNA handles and CBD and therefore need to be near the ends of the protein in order to pull on the entirety of CBD. The construct contains a N-terminal His_6-tag in front of the first cysteine to assist in purification on a nickel column. Protein expression and purification of CBD was verified by SDS-PAGE gel electrophoresis (Fig. 2).

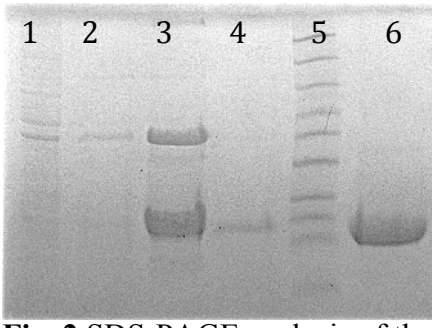

Fig. 2 SDS-PAGE analysis of the protein expression and purification of σ^{54} CBD. This gel shows cell pellet (lane 1), complex mixture loaded on the nickel column (lane 2), flow-through (lane 3), wash (lane 4), ladder (lane 5), and the elution (lane 6).

Denaturation of CBD is necessary before reacting with DTDP because the free cysteines on each end come in close proximity to one another in the tertiary structure. When folded, intramolecular disulfide bond formation happens faster than the cysteines react with DTDP. The pure protein was buffer exchanged out of DTT and into an excess of DTDP on a Sephadex G-25 column by gravity filtration and allowed to react overnight. The addition of the DTDP protecting groups to the two cysteines was verified by mass spectrometry (Fig. 3). Both the reacted and unreacted protein appeared in the mass spectrum indicating that DTDP did react with some CBD molecules others formed intramolecular disulfides and did not react. The reacted protein can be stored at 4 °C for three months.

a **b**

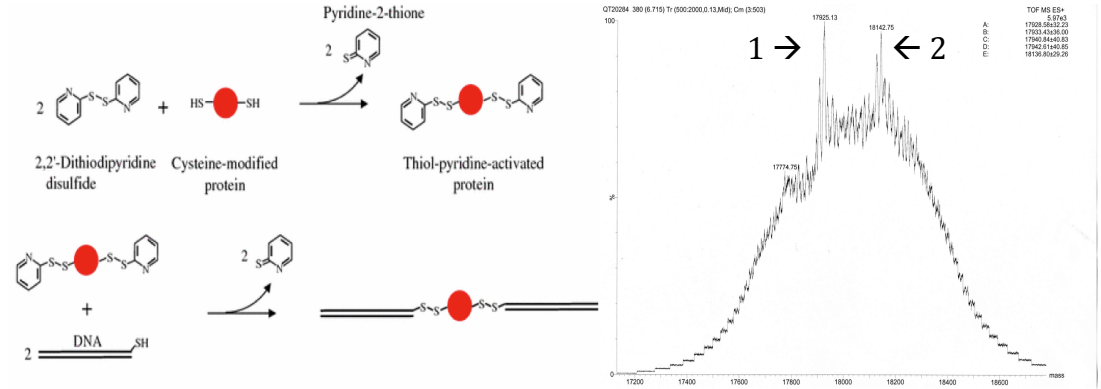

Fig. 3 Attachment of DNA handles to CBD. **a** Schematic of DTDP activation of protein and subsequent handle attachment [9]. **b** Mass Spectrometry analysis of DTDP reaction with CBD. The peak at 1 represents the mass of the unreacted protein. The peak at 2 represents the mass of the reacted protein, containing thiol-pyridine groups covalently bound to the cysteine at each end of the protein.

DNA-handle generation and purification was verified by DNA sequencing and agarose native gel electrophoresis (Fig. 4). The DNA-handles have modified ends that can covalently attach to the

protein on one end and strongly interact with the polystyrene beads on the other. One handle contains a 5' thiol group and a 5' biotin while the other handle contains a 5' thiol group and a 5' digoxigenin. The thiol end reacts with either cysteine on the termini of CBD to form a DNA-protein hybrid. The use of 0.05% tween detergent in this reaction greatly increases the yield. The final result of this attachment is a mixture of DNA-modified proteins, some with both a biotin- and digoxigenin-labeled handle and the rest with unreacted CBD, CBD with only one of the handles, and CBD with two identical handles. This mixture does not require further purification, as the undesired products will not interfere with the molecular tweezers experiment.

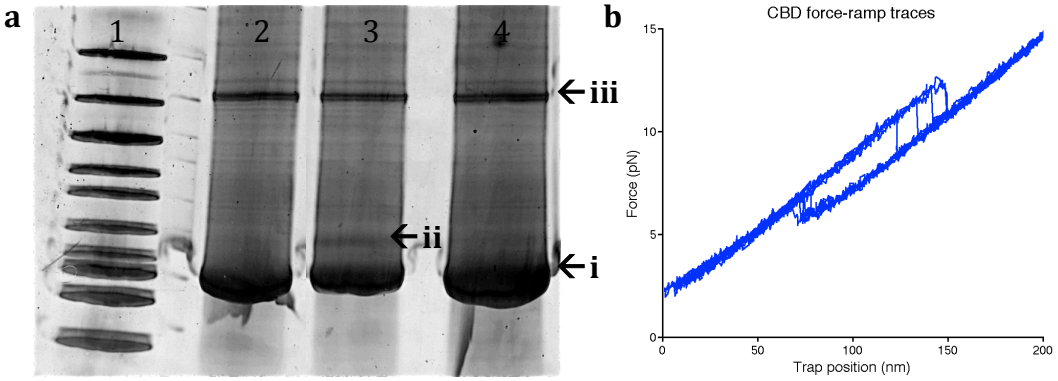

Fig. 4 a Native gel analysis of attachment of DNA-handles to CBD. This DNA gel shows the DNA-handles immediately after the buffer exchange (lane 4), the DNA-handles after reaction with CBD (lane 3), the DNA-handles and CBD reaction after digestion with protease K (lane 2), and the ladder (lane 1). A band showing one DNA-handle connected to the CBD is in lane 3, position ii. A band showing one handle alone is in lane 4, position i, while a band showing two handles covalently bonded to each other is in lane 4, position iii. **b** Force-extension cycles obtained by stretching and relaxing a single molecule of core-binding domain.

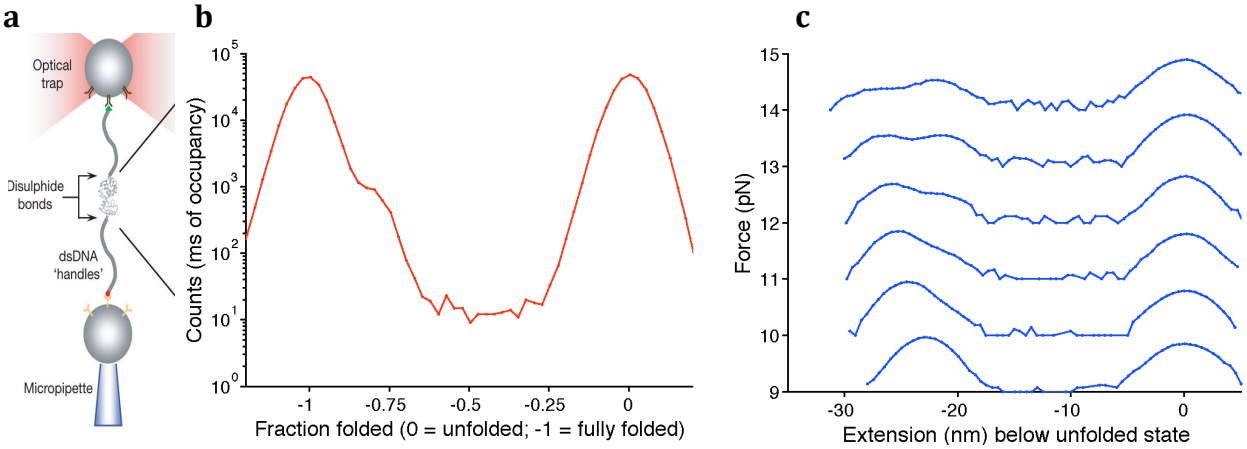

Fig. 5 a Schematic of molecular tweezers experimental set-up [10]. CBD is held by disulfide bonds to the DNA-handles that are attached to the beads. The bead held by the optical trap is coated in anti-dig antibodies that interact strongly with the digoxigenin ends of the DNA-handle. The bead held by the micropipette is coated in streptavidin that interacts strongly with the biotin ends of the other DNA-handle. **b** Amount of time CBD spends in various folded and unfolded

states during force-extension cycles obtained by stretching and relaxing CBD molecules with the molecular tweezers. **c** Force-extension cycles observed by stretching and relaxing CBD at different forces.

Multiple molecules of CBD bound to two unique DNA handles were manipulated using the molecular tweezer set-up depicted in Fig. 5a. Stretching, unfolding, and refolding of CBD was achieved by moving the micropipette relative to the optical trap. The change in momentum of the light beams leaving the trap is measured and can be used to compute the force applied on the molecule [11]. The DNA-handles' contribution to the overall signal is well-characterized and can be separated from that of the tethered protein [11].

The amount of time CBD spends in various folded and unfolded states can be measured by stretching and relaxing CBD molecules over many force-extension cycles causing it to unfold and refold repeatedly. Transitions between completely folded and unfolded states are short-lived compared to the time spent completely folded and completely unfolded. There is no evidence for a fracture point that causes the two subdomains to unfold separately under force. The shoulder of Fig. 5b at around -0.8 fraction folded suggests that 20% of the protein unfolds before the rest of the protein at higher forces, but this is not large enough to be an entire subdomain.

Discussion

NMR studies suggest there may be a conformational fracture point within the CBD that could play a role in transferring an activation event at the N-terminus of σ^{54} to the opening of DNA towards the C-terminus [8]. To test whether this activation mechanism could depend on force, we performed molecular tweezers experiments on CBD of σ^{54} to see if the tugging force of the tweezers would unfold the two CBD subdomains separately. We found that the unfolding of both subdomains of CBD are cooperative, although 20% unfolds separately from the rest of the domain. In Fig. 6, both the blue and the red portion individually represent 20% of the total molecule. The fracture point between the two subdomains does not appear to produce separate unfolding events.

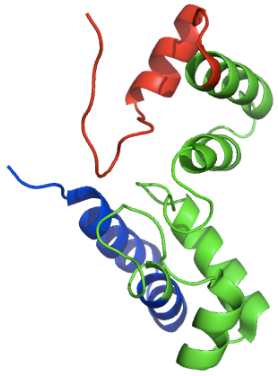

Fig. 6 Structure of CBD showing the two primary candidates for the 20% unfolding event colored blue and red. The first α-helix and half of four-helix bundle on the N-terminus is colored blue and the last α-helix of the three-helix bundle on the C-terminus is colored red.

We propose that the C-terminal helix, represented in red in Fig. 6, accounts for the 20% of the molecule that unfolds separately. It is unlikely that the middle of the molecule would unfold separately because that would greatly disrupt the hydrophobic core that stabilizes the rest of the domain. The N-terminal helix is also less likely because it contains more contacts than the C-terminal helix and 20% of the N-terminus amounts to approximately 1.5 helixes rather than one full helix on the C-terminal end. This leaves the C-terminal helix as the likely cause of the early separate folding event.

References

1. Burgess, R., Travers, A., Dunn, J., Bautz, E. K. (1969). Factor stimulating transcription by RNA polymerase. *Nature*. **221,** 43-46.
2. Mooney, R. A., Darst, S.A., Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell*. **20,** 335-345.
3. Fisher, M.A., Grimm, D., Henion, A.K., Elias, A.F., Stewart, P.E., Rosa, P.A., Gherardini, F.C. (205). *Borrelia burgdorferi* sigma54 is required for mammalian infection and vector transmission but not for tick colonization. *Proc. Natl. Acad. Sci.* USA, **102**, 5162-5167.
4. Saldias, M.S., Lamothe, J., Wu, R., Valvano, M.A. (2008). *Burkholderia cenocepacia* requires the RpoN sigma facto for biofilm formation and intracellular trafficking within macrophages. *Infect. Immun.* **76,** 1059-1067.
5. Hirschman, J., Wong, P.K., Sei, K. Keener, J., Kustu, S. (1985). Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc. Natl Acad. Sci.* USA. **82,** 7525-7529.
6. Wedel, A., Kustu, S. (1995). The bacterial enhancer-binding protein NTRC is a molecular machine: ATP hydrolysis is coupled to transcriptional activation. *Genes Dev.* **9,** 2042-2052.
7. Barrios, H., Valderrama, B., Morett, E. (1999). Compilation and anaylsis of sigma54 dependent promoter sequences. *Nuc. Ac. Res.* **27,** 4305-4313.
8. Hong, E., Doucleff, M., Wemmer, D.E. (2009). Structure of the RNA polymerase core-binding domain of sigma54 reveals a likely conformational fracture point. *Journal of molecular biology.* **390,** 70-82.
9. Cecconi, C., Shank, E., Marquesee, S., Bustamante, C. (2008). Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J.* **37,** 729-738.
10. Shank, E., Cecconi, C. Dill, J., Marqusee, S., Bustamante, C. (2010). The folding cooperativity of a protein is controlled by its chain topology. *Nature*. **465**, 637-640.
11. Smith, S., Cui, Y., Bustamante, C. (2003). Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* **361,** 134-162.