Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2017

Structural and Functional Studies of Mycobacterial General Transcription Factors RbpA and CarD

Elizabeth Hubin

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Recommended Citation
Hubin, Elizabeth, "Structural and Functional Studies of Mycobacterial General Transcription Factors RbpA and CarD" (2017). *Student Theses and Dissertations*. 407.
http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/407

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.

STRUCTURAL AND FUNCTIONAL STUDIES OF MYCOBACTERIAL GENERAL TRANSCRIPTION FACTORS RbpA AND CarD

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Elizabeth Hubin
June 2017

© Copyright by Elizabeth Hubin 2017

STRUCTURAL AND FUNCTIONAL STUDIES OF MYCOBACTERIAL GENERAL TRANSCRIPTION FACTORS RbpA AND CarD

Elizabeth Hubin, Ph.D.

The Rockefeller University 2017

Gene expression in bacteria is highly regulated at the step of transcription initiation. For decades, the vast majority of biochemical and kinetic studies have used *Escherichia coli* (*Eco)* RNA polymerase (RNAP) as a model for studying prokaryotic transcription initiation. However, properties of *Eco* RNAP are not representative of RNAPs from other bacterial species. Transcription in mycobacteria is distinct from *Eco* by its formation of unstable initiation complexes and its dependence on two transcription factors, RbpA and CarD, which are essential in the pathogen *Mycobacteria tuberculosis* (*Mtb*).

In this thesis, I report the structural and biochemical characterization of RbpA and describe how it works synergistically with CarD to compensate for the unstable RNAP-promoter complexes formed in mycobacteria. I describe two crystal structures, *Mtb* RbpA-SID/σ^{A}_{2} complex and RbpA bound to a *Mycobacterium smegmatis* (*Msm*) RNAP initiation complex, which reveal the structural mechanism of RbpA. The C-terminal σ-interacting domain (SID) of RbpA binds to domain 2 of σ (σ^{A}_{2}), while RbpA's core domain interacts with the β' subunit of RNAP. A basic linker (BL) connecting these two

domains contacts promoter DNA upstream of the transcription bubble. Functional and kinetic analyses reveal that the interaction between the RbpA-BL and promoter DNA is the basis for transcription activation by RbpA.

A synergistic effect between RbpA and CarD, striking in both transcription assays and in a kinetic assay measuring RPo formation, indicates that these two factors work together to activate transcription. By teasing apart the effects of the two factors on kinetic steps in RPo formation, this work shows that the RbpA and CarD cooperatively drive formation of a kinetic intermediate species during RNAP open complex formation.

The work in this thesis provides a structural and functional understanding of transcription stimulation by RbpA and CarD, which we hypothesize make up part of the general transcription machinery in mycobacteria. There are currently no reported structures of mycobacterial RNAP, so the structure of the *Msm* RNAP initiation complex presented in this thesis will provide a platform for future studies studying the mycobacterial transcription system.

Acknowledgements:

Neither my research nor this thesis would be possible without the many individuals who mentored, challenged, and supported me during my years as a PhD student at The Rockefeller University.

I want to convey my deepest gratitude to Dr. Seth Darst. His scholarship and intellectual commitment to science inquiry have been both inspirational and instrumental in my quest to be a careful, critical, and questioning scientist. I especially appreciate how he patiently taught me the theoretical and technical aspects of crystallography and kinetics.

Dr. Elizabeth Campbell could not have been a better mentor throughout my time in the Darst Lab. She not only spearheaded and guided this project, but she also made herself available whenever I needed consultation, advice, and encouragement. I will always appreciate the time and interest she invested in me, providing an impressive role model as a strong woman in science. I appreciate, too, the trust that Liz and Seth showed in letting me mentor students in our Lab. The opportunities to work with Joshua Flack, Abbas Nazir, Catherine Xu, and Natalia Munoz allowed me to test and advance my own abilities to teach and mentor. I also want to thank these committed students for their contributions to this project.

The research summarized in this thesis has relied on the intellectual and experimental contributions of colleagues and collaborators whose work shaped the progression of our project. In particular, I want to acknowledge Mark Paget's Lab at the University of Sussex for initiating our collaboration on the RbpA project and assisting

with *in vivo* and biochemical experiments, and also Mike Glickman's Lab at Memorial Sloan Kettering Cancer Center for performing *Msm in vivo* experiments.

I am grateful to my committee members Dr. Charles Rice and Dr. Luciano Marraffini for providing on-going feedback and encouragement as I shared the stages of my research. Their enthusiastic interest in my project has been both highly informative and motivational.

I cherish that many working relationships at the Rockefeller University developed into close friendships. Thank you, Beth Davis, for your supportive and encouraging presence, and Nathaniel Braffman, for not only your helpful contributions to this project but also your knack for making lab a contagiously fun place to be serious.

I will always appreciate that Jeanne Garbarino saw some potential in me when she invited me to join the Science Outreach Program, and I am grateful that she then became an inspiring pedagogical mentor to me. Our instructional adventures sparked my passion for teaching science.

The Deans Office deserves special recognition for offering much needed support as I pulled the pieces of my research into a doctoral thesis. A special thanks to Dean Syd Strickland and Assistant Dean Emily Harms who have made it an enjoyable privilege to be a graduate student at The Rockefeller University.

My heartfelt thanks to Ari Halper-Stromberg, especially for his support and uncanny intuition—knowing when I needed solitary concentration and when, instead, I could benefit from memorable distractions. And to my parents, thank you for the unwavering love, guidance, and confidence you've provided during what has truly been a remarkable journey of intellectual and personal growth.

Table of Contents:

List of Figures:

List of Tables

List of Abbreviations:

Bsu	*Bacillus subtilis*
CD	Core domain
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
Eco	*Esherichia coli*
GTF	General transcription factors
LSI	Lineage specific insert
Mbo	*Mycobacterium bovis*
Msm	*Mycobacterium smegmatis*
MTC	Mycobacterium tuberculosis complex
Mtb	*Mycobacterium tuberculosis*
NMR	Nuclear Magnetic Resonance
nt	nucleotide
NTD	N-terminal domain
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RID	RNAP-interacting domain
RNA	Ribonuceleic acid
RNAP	RNA Polymerase
RPc	RNAP closed complex
RPo	RNAP open complex

rRNA	ribosomal RNA
Sco	*Streptomyces coelicolor*
SDS	Sodium dodecyl sulfate
SID	Sigma-interacting domain
Taq	*Thermus aquaticus*
TB	Tuberculosis
TEC	Transcription elongation complex
Tth	*Thermus thermophilus*
ZBD	Zinc-binding domain

Chapter 1: Introduction

1.1 Mycobacterium tuberculosis

The bacterial pathogen *Mycobacterium tuberculosis* (*Mtb)* is the causative agent of tuberculosis (TB), a serious infection in the lungs and an on-going world health problem responsible for approximately 1.5 million deaths per year worldwide (Center for Disease Control, 2015 data). *Mtb* has co-existed with humans for more than 40,000 years (Wirth et al., 2008) and has evolved to persist latently in a human host for decades. *Mtb* has adapted to multiple immune evasion strategies, including the ability to replicate within host macrophage cells (de Chastellier, 2009). An estimated 30% of the world's population is infected with latent *Mtb* (CDC, 2015 data), which presents no clinical manifestations; however, reactivation of latent *Mtb* occurs in a subset of individuals, generally in those who are elderly or immunocompromised, which can cause severe illness and mortality. Further, at least 5% of TB cases across the globe are drug resistant (CDC, 2015), which has exacerbated the health crisis, highlighting the need for novel therapeutic strategies for tuberculosis infection.

In addition to *Mtb,* there are other tuberculosis-causing mycobacteria, genetically very similar to *Mtb*—including *Mycobacterium bovis (Mbo),* used in many of the studies described in this thesis. As a group, these bacteria are defined as the mycobacterium tuberculosis complex (MTC). In comparison to *Escherichia coli (E. coli* or *Eco*), which has a doubling time of 20-30 minutes, MTC bacteria are very slow growing, with a doubling time of 14-24 hours. The slow growth of MTC bacteria makes *in vivo* studies of the organisms difficult. Thus *Mycobacteria smegmatis (Msm)*, with its doubling time of ~3 hours, is often used as a model organism to study MTC bacteria.

Through fine control of stress response genes, *Mtb* has adapted to survive an onslaught of attacks from the host immune response (Flentie et al., 2016). Transcription initiation is a major point of gene regulation in bacteria, and thus an understanding of the *Mtb* transcription system would provide insight into essential regulatory processes unique to an important pathogen.

1.2 General mechanism of bacterial transcription

Transcription, the process in which information encoded as DNA is copied into messenger RNA (mRNA), is the first step of gene expression and subsequently the primary point of gene control in prokaryotes. The transcription cycle can be divided into three phases: initiation, elongation and termination (**Figure 1.1)**. Transcription is performed by the evolutionarily conserved ~400 kDa catalytic core RNA Polymerase (RNAP) enzyme (subunit composition $\alpha_2\beta\beta'\omega$) (Burgess, 1969; Burgess and Travers, 1970). Core RNAP interacts non-specifically with DNA and is active for transcription elongation, but requires the promoter-specificity subunit, σ, for promoter-specific initiation of transcription (Burgess et al., 1969; Murakami and Darst, 2003). During initiation, σ associates with core RNAP to form the holoenzyme, which is directed to transcription start sites through sequence-specific interactions made between σ and promoter DNA. Upon binding, σ facilitates melting of the DNA to form a transcription bubble, extending from the -11 to the +2 base (relative to the +1 start site), yielding the transcriptionally competent open complex (RPo) (Bae et al., 2015; Murakami et al., 2002b).

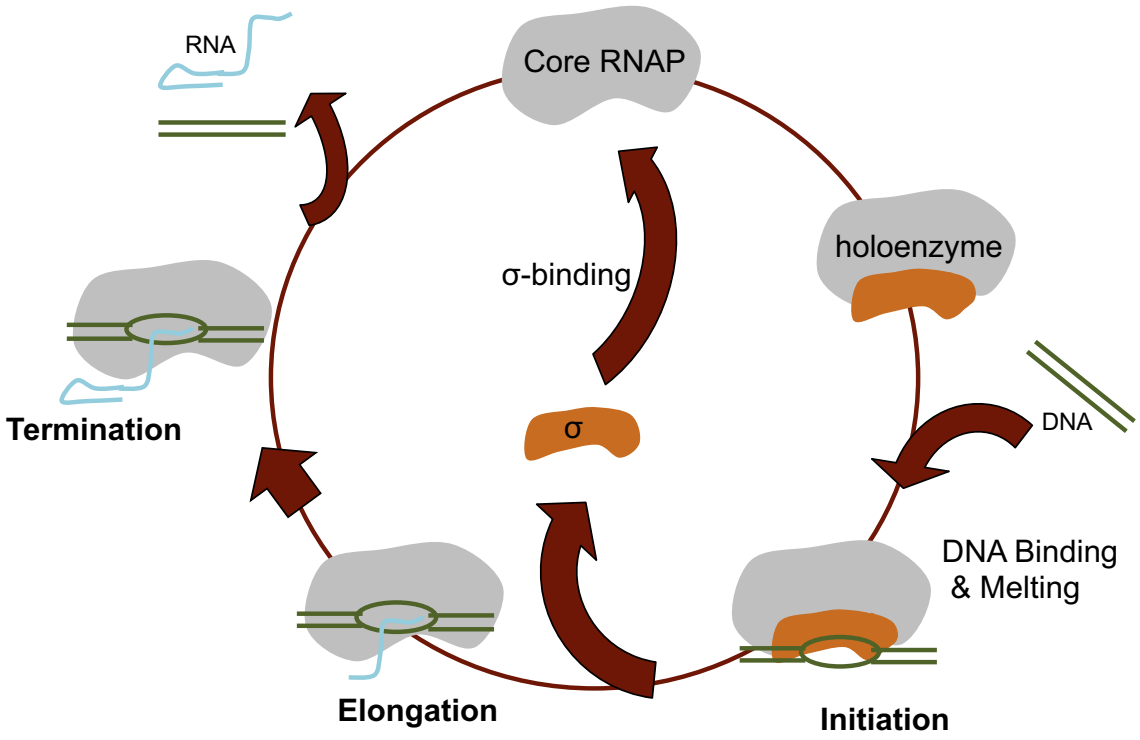


Figure 1.1: Transcription cycle in bacteria. Transcription initiation begins when core RNAP binds σ, forming the holoenzyme, which is competent for sequence specific binding of promoter DNA. σ facilitates melting of promoter DNA to form a transcription bubble (from the -11 to the +1 start site), yielding the transcriptionally active RNAP open complex (RPo). In the presence of NTP substrate, the RNAP initiation complex will repeatedly produce RNA transcripts until it produces a transcript long enough to disrupt interactions between core RNAP and σ. As the RNAP enters into productive elongation, σ is stochastically released. Elongation continues until the RNAP encounters a termination signal, at which point the RNAP releases the mRNA transcript and DNA. Core RNAP and σ are recycled and can begin the pathway again.

In the presence of nucleotide substrate, the RPo begins RNA synthesis. Ribonucleotide (rNTPs) bases pair with template strand DNA, and RNAP catalyzes formation of the phosphodiester bond. Before entering transcription elongation, the RNAP holoenzyme reiteratively produces short RNA transcripts (2-12 nucleotides in length) until it produces a transcript long enough to destabilize interactions between core RNAP and σ (Murakami and Darst, 2003). Upon release of σ, the enzyme escapes the promoter and enters into elongation. The RNAP transcription elongation complex (TEC) continues until it encounters a termination signal (Washburn and Gottesman, 2015), at which point RNAP releases the RNA transcript and dissociates from the DNA.

1.3 Promoter recognition by RNAP

In the bacterial cell, σ factors are key regulators of transcription initiation that function by directing core to the appropriate transcription start sites. All bacteria contain a single housekeeping (primary) σ factor that directs the bulk of transcription during exponential growth (σ^{70} in *Eco*, σ^{A} in *Taq* and mycobacteria) (Gruber and Gross, 2003; Murakami et al., 2002b).

Primary σ factors are composed of three highly conserved domains (σ_2, σ_3, σ_4) connected by flexible linkers, as well as a poorly conserved (although characteristically acidic), flexible N-terminal region ($\sigma_{1.1}$). Most of these σs also have a non-conserved region (NCR) inserted within σ_2 between regions 1.2 and 2 that vary significantly in size, sequence, and structure between bacterial lineages.

The current paradigm for recognition of housekeeping genes is that primary σ factors recognize promoter DNA through interactions with DNA upstream of the +1 start

site (**Figure 1.2**). The -10 promoter element (or Pribnow box, consensus: TATAAT) (Pribnow, 1975) and the -35 promoter element (consensus: TTGACA) are recognized by σ_2 and σ_4, respectively. A 16-19 bp spacer region, with no consensus sequence but a consensus length of 17 bp, separates the -10 and -35 elements.

Subsets of housekeeping promoters contain an "extended -10," a TG at the -14 and -13 positions, which is contacted by σ_3. In *Eco,* promoters with weak or lacking -35 elements may depend on an extended -10 to help facilitate promoter binding and transcription initiation (Keilty and Rosenberg, 1987; Mitchell et al., 2003). The region between the -6 and -3, referred to as the discriminator, is recognized by σ_2 upon DNA strand separation. The discriminator region is an important determinant of stability of the RNAP-promoter complex (Haugen et al., 2006).

In addition to promoter elements recognized by σ, some bacteria, including *Eco,* have promoters that contain an upstream sequence recognized by the C-terminal domains of the α subunit (αCTD), attached to the RNAP enzymes by flexible linkers. These A/T rich sequences, approximately 40-60 bases upstream of the -35, are referred to as UP elements. All of these DNA elements—the -35 element, -10 element, extended-10, UP-element, and discriminator element—can modulate the strength of a promoter, affecting both the kinetics of initial RNAP binding and RPo formation (Haugen et al., 2008; Ruff et al., 2015).

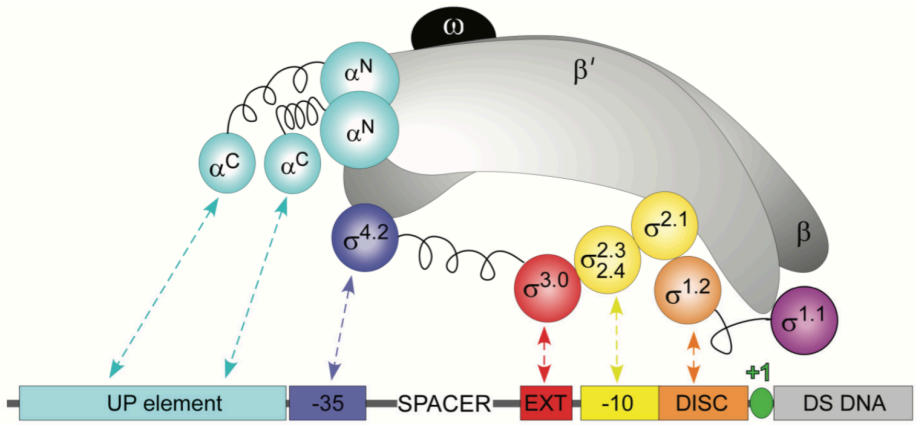


Figure 1.2: Recognition of housekeeping promoters by RNAP. Promoter elements involved in sequence specific interactions are highlighted and color-coded corresponding to the region of the holoenzyme by which they are recognized. Up element is colored cyan, -35 element is blue, the extended -10 is red, the -10 element is yellow, and the discriminator is orange. Flexible linkers that are connecting σ domains and the α-CTDs to core are shown as springs. Figure is adapted from Ruff et al., 2015.

Most bacteria also contain alternate σ factors that are expressed during stationary phase or stress conditions and drive transcription of stress-response genes by recognizing alternative promoter sequences. Group 2 sigma factors are similar in sequence and structure to primary σ factors (group 1) but are not required for viability, and they sometimes lack region 1.1 and the NCR (Gruber and Gross, 2003). Group 3 σ factors are more divergent and lack region 1.1 and a NCR, and group 4 σ factors (also called ECF factors) are the most divergent, only containing σ_2 and σ_4 (Feklistov et al., 2014). The number of alternate σ factors in the cell varies in different bacteria species—*Eco* has 7 (Osterberg et al., 2011), *Mtb* has 12 (Flentie et al., 2016), while *Streptomyces coelicolor* (*Sco*) has as many as 65 (Bentley et al., 2002).

1. 4 Structure of prokaryotic RNAP

Crystal structures of RNAPs from thermophilic bacteria (*Thermus aquaticus* (*Taq*) and *Thermus thermophilus* (*Tth*) (Bae et al., 2015b; Murakami et al., 2002b; 2002a; Vassylyev et al., 2002; Zhang et al., 1999) and *Escherichia coli* (*Eco*) (Bae et al., 2013; Murakami, 2013; Zuo and Steitz, 2015) have enabled a better understanding of the process of transcription in bacteria. RNAP structures reveal that the overall shape of the core enzyme resembles a crab claw, with the β and β' subunits each forming a pincer (**Figure 1.3a**). The active site contains a catalytic Mg^{2+} ion and is buried in the main channel defined by the β, β' interface (Zhang et al., 1999). The main channel (~25 Å in diameter) accommodates the template strand DNA and the DNA-RNA hybrid at the growing strand of the RNA chain, while a narrower secondary channel (10-12 Å) allows

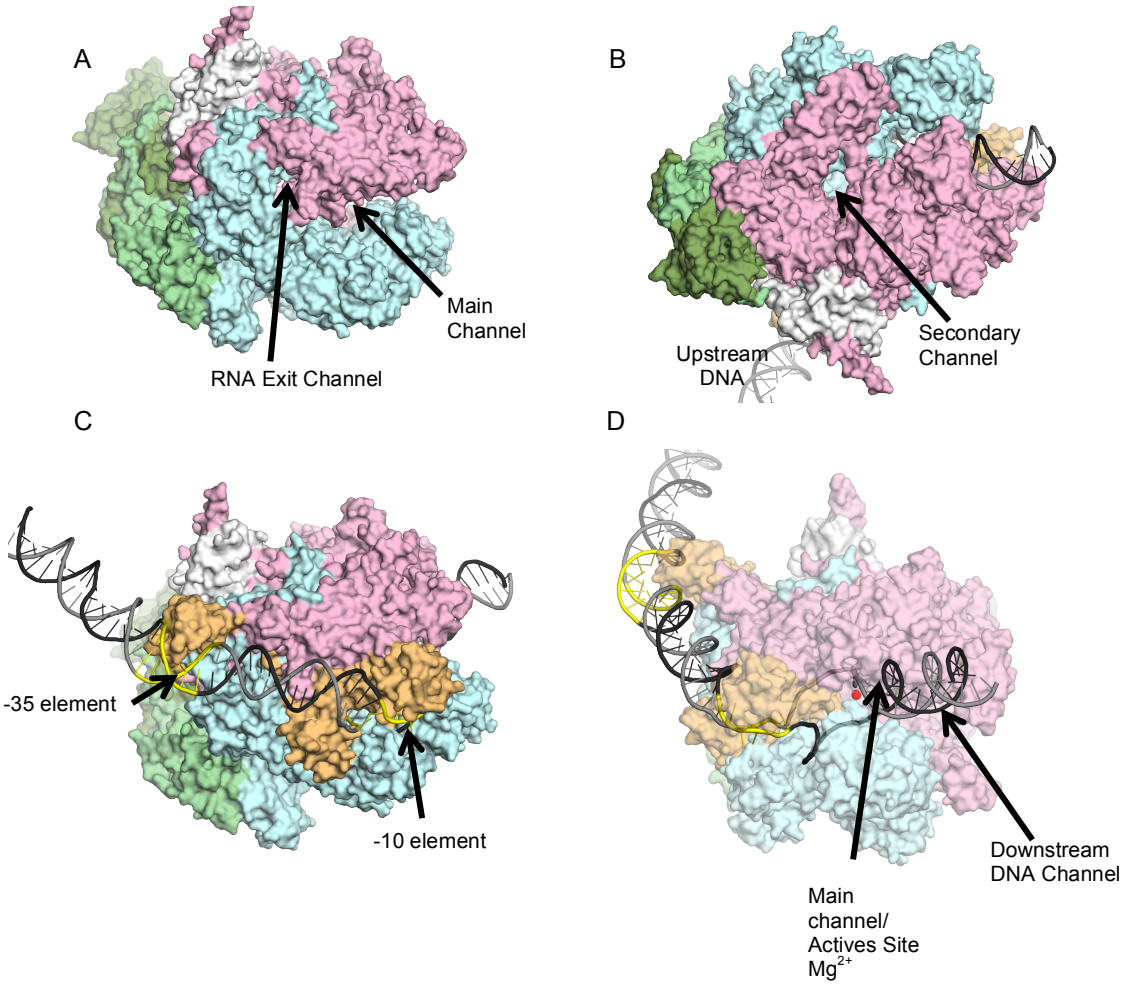


Figure 1.3: Structure of prokaryotic RNAP. Structural models of RNAP generated from crystal structures of *Taq* RNAP (1KU2, Bae et al., 2015). β is shown in cyan, β' is pink, α1 and α2 are green, ω is white, and σ is orange. Template strand DNA is light gray and non-template strand DNA is dark gray. -10 and -35 elements are shown in yellow. Catalytic Mg^{2+} is shown in red (A) RNAP core (B-D) initiation complexes.

free nucleotide substrate to enter into the active site (Zhang et al., 1999) (**Figure 1.3b-d)**. During elongation, nascently transcribed RNA is expelled through an RNA exit channel (Murakami and Darst, 2003). Unbound σ in solution does not bind promoter DNA as it is compactly folded so DNA binding domains are inaccessible (Schwartz et al., 2008). However, in the context of the holoenzyme, the DNA binding determinants of σ are revealed and spatially positioned to interact with promoter elements (**Figure 1.3c**). Holoenzyme structures reveal that each σ domain, as well as the linkers, makes extensive interactions with core, forming an interface >8000 Å (Murakami et al., 2002a).

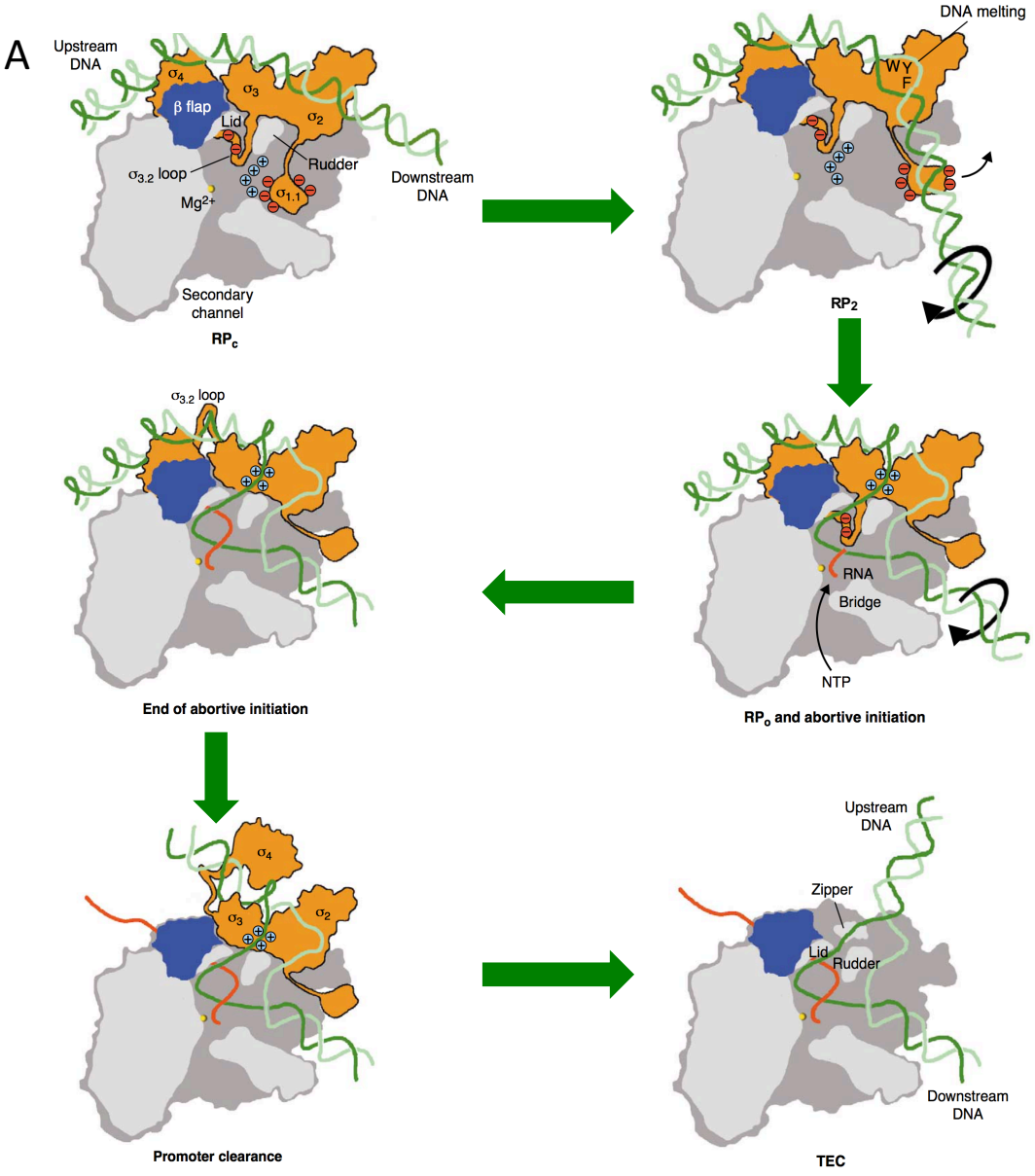
To date, all structures of bacterial RNAP have been from *Eco* or *Thermus* species. Although many core RNAP subunits are highly conserved in sequence across all bacteria phyla, the importance of determining RNAP from a wide range of bacteria is highlighted by: 1) the existence of RNAP structural inserts specific to certain bacterial lineages that have been proposed to modulate the activity of the enzyme, either directly or by interacting with cis-acting regulators (Lane and Darst, 2010b; 2010a); 2) the existence of essential transcriptional regulators that do not have homologues in *Eco* or *Taq* (Bao et al., 2012; Paget et al., 2001; Stallings et al., 2009); 3) biochemical studies that reveal significant functional differences in RNAP from various species that may be highlighted in a structural analysis (Davis et al., 2015; Schroeder and deHaseth, 2005; Whipple and Sonenshein, 1992); and 4) the study of small molecule drugs that do not inhibit *Eco* or *Taq*.

1.5 Structural and kinetic mechanism of transcription initiation in bacteria

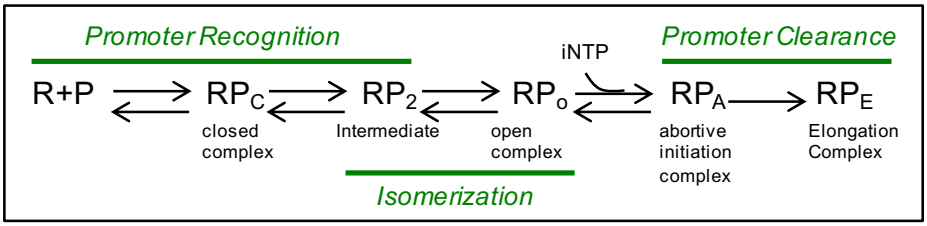
RPo is a multi-step pathway during which the initiation complex undergoes large conformational changes and transitions through discrete intermediate states (**Figure 1.4**) (Saecker et al., 2011). Kinetic studies with *Eco* RNAP on Eco promoter *lac*UV5 and T7A1 have been able to characterize three RNAP-promoter populations: an initial "closed complex" intermediate (RPc); a second intermediate state (RP2 or RPi on *lac*UV5 and T7A1); and then a final open complex (RPo) (Buc and McClure, 1985). However, it is likely that additional intermediates occupy this pathway. On the promoter λPR, for instance, three intermediates have been identified—*I1*, *I2*, and *I3* (Roe et al., 1984; Saecker et al., 2002).

Intermediate states are unstable and transient and are therefore challenging to characterize structurally; thus, structural models of RNAP-promoter intermediates have been inferred by combining structural information with real-time DNAse and hydroxyradical footprinting assays, which map DNA-protein interactions and DNA distortions (Craig et al., 1998; Davis et al., 2007; Gries et al., 2010; Rutherford et al., 2009; Saecker et al., 2002; Sclavi et al., 2005). Initial binding of RNAP to DNA results in the RPc intermediate in which DNA is in duplex form, and downstream DNA is unbent. RPc formation is thought to be mediated by holo-DNA interactions upstream of the -10 element, including the binding of σ_4 to the -35, the binding of σ_3 to the extended -10, and the binding of the αCTD to the UP-element.

Figure 1.4 (Following page): Overview of structural changes during transcription initiation. (A) The holoenzyme forms a closed complex (RPc) upon initial binding to DNA. An intermediate (RP2) forms upon DNA bending and initial melting of the -10. RPo forms when the DNA is melted, and the template strand is in the active site channel. NTPs enter the active site through the secondary channel, and RNAP undergoes abortive initiation until a transcript is long enough to dislodge σ region 3.2 from the RNA exit channel, breaking interactions between σ_4 and σ_3 and core RNAP. RNAP can then clear the promoter, σ is stochastically lost, and RNAP enters elongation. Adapted from Murakami and Darst, 2003. (B) Kinetic schematic of transcription initiation.


A
Upstream DNA
σ4
σ3
σ2
β flap
Lid
σ3.2 loop
Mg2+
Rudder
σ1.1
Downstream DNA
Secondary channel
RPc
DNA melting
WY F
RP2
σ3.2 loop
End of abortive initiation
RNA
Bridge
NTP
RPo and abortive initiation
σ4
σ3
σ2
Promoter clearance
Upstream DNA
Zipper
Lid
Rudder
Downstream DNA
TEC

B


Promoter Recognition
Promoter Clearance
iNTP
R+P
RPC
RP2
RPo
RPA
RPE
closed complex
Intermediate
open complex
abortive initiation complex
Elongation Complex
Isomerization

Sequence-specific recognition of the -10 element is coupled with melting as σ_2 recognizes the flipped out -11A base, facilitating a 90 degree bend of the downstream DNA (Feklistov and Darst, 2011). Subsequent interactions between σ_2 and promoter DNA result in an intermediate (RP2) in which the -10 element is melted and the bend at the -10 element places the downstream DNA, which is still duplex, across the entrance to the ββ' main channel (Murakami and Darst, 2003). Final open complex (RPo) occurs upon formation of the full transcription bubble and accommodation of the downstream DNA (single-stranded and duplex) into the appropriate sites within the main channel. At this point the transcription bubble is fully formed (extending from the -11 to the +2 base), and downstream duplex DNA from the +5 to the +12 position is clamped by the β and β' subunits.

Upon formation of RPo, and as rNTPs diffuse through the secondary channel into the active site, RNAP begins to catalyze the synthesis of RNA. For RNAP to successfully enter into the elongation phase, it must break the strong contacts between σ and the promoter. Prior to promoter escape, the enzyme undergoes a process called abortive initiation (Munson and Reznikoff, 1981; Vo et al., 2003) where a loop in σ_3 ($\sigma_{3.2}$) blocks the path of elongating RNA, so the enzyme continually produces short RNA transcripts that are released, likely through the secondary channel (Murakami and Darst, 2003; Vassylyev et al., 2002). Once RNAP is able to produce a transcript of 15-16 nucleotides (Goldman et al., 2009), $\sigma_{3.2}$ is spatially dislodged by the RNA, and contacts between core and $\sigma_{2\text{-}3}$ are weakened, allowing for RNAP to enter into elongation. After promoter clearance, σ is stochastically lost from the transcription elongation complex (TEC)

(Shimamoto et al., 1986), but some population of holoenzyme is retained via the interaction between the flap and σ_4 (Mooney et al., 2005).

Region 1.1 of σ also plays a role in RPo formation. An *Eco* holoenzyme structure without DNA shows that $\sigma_{1.1}$ occupies the downstream DNA channel (Bae et al., 2013). Placement of DNA into the channel during RPo formation must therefore be accompanied by displacement of $\sigma_{1.1}$.

1.6 Pausing during transcription initiation

During elongation, RNAP is prone to pausing on the promoter. Pausing serves as a mechanism for transcription regulation, allowing for coordination of transcription and translation and providing an opportunity for transcription regulators to bind to the RNAP (Landick, 2006). Mechanisms for pausing are varied and can involve sequence specific interactions between RNAP and nucleic acid strands (both DNA and RNA) or with factors that interact with the TEC. Pausing can also occur when σ is bound to a TEC (either by remaining bound to RNAP post initiation or by re-binding the elongation complex), and σ interacts with promoter sequences resembling a -10 element (Brodolin et al., 2004; Mooney and Landick, 2003). Thus σ-dependent pausing also represents a mechanism of gene control employed by bacteria.

1.7 Mycobacteria RNAP forms unstable promoter complexes

For decades, the vast majority of biochemical and kinetic studies have used *Eco* RNAP as a model for studying transcription initiation. However, recent studies have shown that the properties of *Eco* RNAP are not representative of RNAPs from other

bacterial species (Davis et al., 2015; Rammohan et al., 2015; Schroeder and deHaseth, 2005; Whipple and Sonenshein, 1992). Recent work studying the activity of *Eco* and *Mbo* RNAPs on the same promoters has shown that *Eco* RNAP forms essentially irreversible open complexes while *Mbo* RNAP forms an unstable open complex with a half-life of a few minutes or less (**Figure 1.5**) (Davis et al., 2015). Further, *Mbo* RNAP was found to be significantly more sensitive to chloride salt than *Eco* RNAP. Despite the overall weaker activity of *Mbo* compared to *Eco* and the marked difference in RPo stability, DNA footprinting assays mapping the DNA-protein interactions and DNA distortions indicate that the RNAP/promoter physical interactions are similar between the two enzymes. KMnO4 assays, which probe promoter melting, indicate that the two RNAPs generate the same transcription bubbles and participate in roughly the same protein/DNA interactions on the promoters studied.

The mycobacterial transcription system is also distinct from the *Eco* system in its deviation from classical promoter architecture. Genome wide analysis reveals that although mycobacteria do have a conserved -10 element (TAnnnT) similar to *Eco,* they lack a consensus -35 element, and the majority of promoters have sequences in the -35 region position that bear no resemblance to the TTGACA consensus sequence in *Eco* (Agarwal and Tyagi, 2006; Bashyam et al., 1996; Cortes et al., 2013; Newton-Foot and Gey van Pittius, 2013). Further, only a small percentage of TAnnnT mycobacterial promoters contain an extended -10 TGn motif (Cortes et al., 2013), which are commonly present in *Eco* promoters lacking a -35.

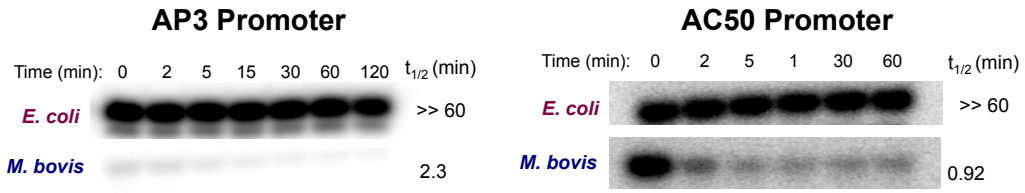


Figure 1.5: *Mbo* forms unstable promoter complexes relative to *Eco.* Abortive initiation half-life assays, in which RPo is challenged by competitor, show that on both the *rrnAP3* promoter (AP3) and synthetic AC50 promoter, *Eco* forms an irreversible RPo while *Mbo* has a half-life of a couple minutes or less. Adapted from Davis et al., 2014.

Additionally, of the mycobacterial promoters that have been described in literature, there is a noticeable absence of classical A/T rich UP elements that would form sequence-specific interactions with the RNAP-αCTD (Arnvig, 2005); however, no genome-wide analyses of this region have been reported, so further studies are necessary to determine whether a subset of mycobacterial promoters do contain UP-elements.

The -10 element seems to be the dominant determinant of sequence-specific promoter recognition for mycobacterial RNAP, and it is possible that the G/C rich nature of the bacterial precludes it from having many of the classical promoter architecture found in *Eco*. Generally, mycobacterial promoters, even those that contain a TTGACA-like -35 element, have been described as weak compared to those in *Eco.* It has been proposed that weaker promoters may be advantageous to the pathogenic nature of *Mtb* (Bashyam et al., 1996) because the bacteria can remain dormant and evade the host immune system. Certainly weaker RNAP enzymes combined with weaker promoters could contribute to the slow growth of MTC bacteria.

1.8 Transcription factor CarD stabilizes mycobacterial RPo

The weak activity of *Mbo* RNAP is, to some extent, rescued by the transcription factor CarD, an essential transcription factor in *Mtb* that is not present in *Eco* (Stallings et al., 2009). CarD was found to extend the half-life of *Mbo* RNAP on the endogenous rRNA promoter *rrnAP3* (AP3*)* more than 10-fold (**Figure 1.6**) (Davis et al., 2015). CarD activates transcription in a manner non-canonical to the classical model of transcription activation established through the studies of activators in *Eco* (Bae et al., 2015a; Srivastava et al., 2013).

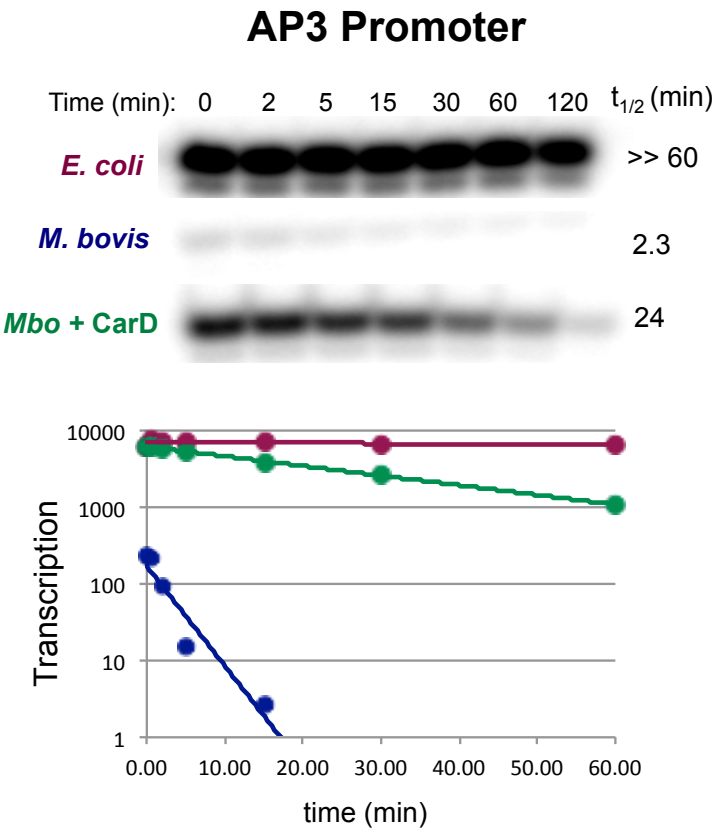


Figure 1.6: CarD stabilizes RPo. In abortive initiation half-life assays, CarD extends the half-life of *Mbo* RNAP open complex on AP3 promoter more than 10 fold. Adapted from Davis et al., 2014.

In the classical paradigm, sequence-specific DNA-binding proteins recruit RNAP to promoter DNA through interactions with σ_4, the αCTD, or both (Browning and Busby, 2004). In contrast, CarD binds the β1' lobe of RNAP through its N-terminal RNAP interacting domain (RID) (Srivastava et al., 2013; Stallings et al., 2009). A crystal structure of the *Thermus* RNAP RPo bound to *Tth* CarD shows that the helical C-terminal domain (CTD) of CarD interacts with the non-template strand DNA upstream of the transcription bubble (-14-10) and wedges a conserved Trp residue, W86 (W85 in *Mtb*), into the splayed minor groove at the -12 nt position (**Figure 1.7**) (Bae et al., 2015a), providing a structural mechanism for how CarD physically prevents transcription bubble collapse.

CarD is widely distributed among bacteria species (Bae et al., 2014); however, it is notably absent in *Eco* but present in *Bacillus* and *Thermus*, which like mycobacteria, have been shown to have less-stable promoter complexes than *Eco* (Schroeder and deHaseth, 2005; Whipple and Sonenshein, 1992). In *Msm,* CarD has been shown to be a global regulator of transcription, present at essentially all promoter regions (Landick et al., 2014; Srivastava et al., 2013). Thus CarD can be thought of as a general transcription factor, likely always present on the mycobacterial RNAP initiation complex, similar to an additional subunit.

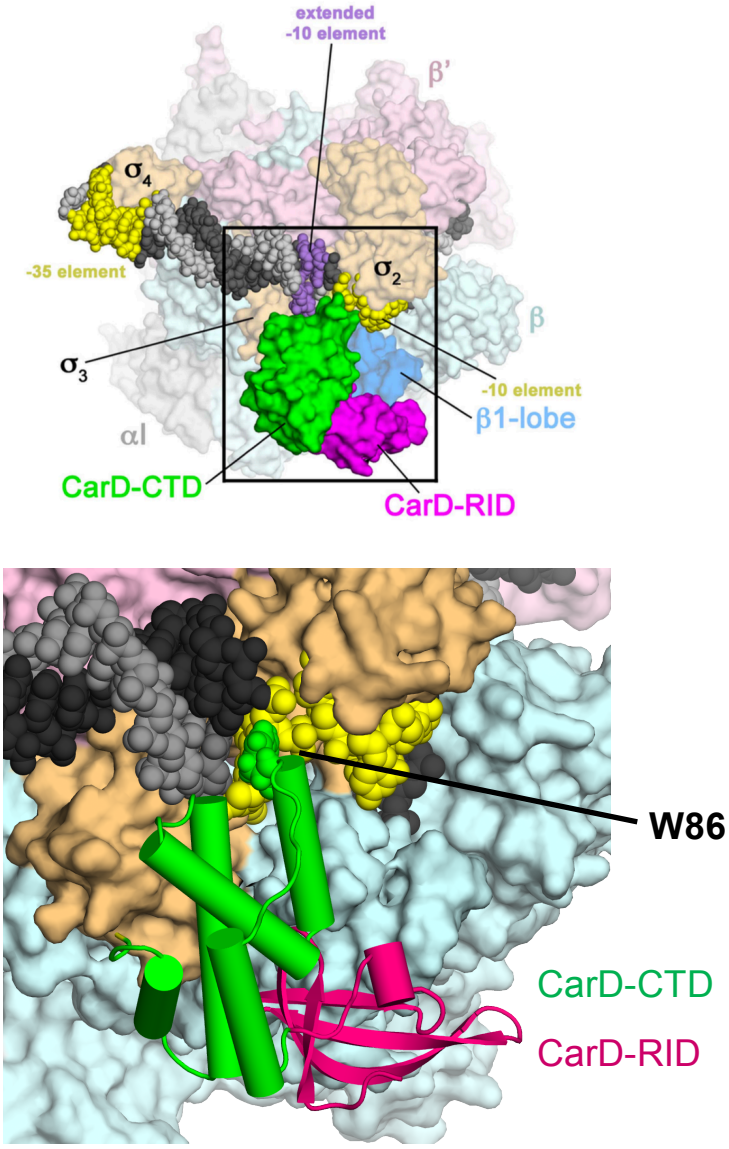


Figure 1.7: CarD uses a minor groove mechanism to stabilize RPo. A crystal structure of *Taq* intiation complex bound to *Thermus* CarD reveals the mechanism for RPo stabilization by CarD. Top: RPo bound to CarD. Bottom: CarD wedges a conserved Trp residue into a splayed minor grove at the upstream edge of the transcription bubble, preventing bubble collapse. Adapted from Bae et al., 2015.

1.8 RbpA is an essential transcription factor in *Mtb*

Like CarD, the Actinobacterial-specific transcription activator RbpA is essential for growth in *Mtb* (Forti, et al., 20011). RbpA was originally discovered in *Sco*, where it was identified to be a major component of RNAP holoenzyme (Paget et al., 2001). RbpA has been shown to specifically activate transcription by holoenzyme complexes containing group 1 or group 2 σ factors (Bortoluzzi et al., 2013; Hu et al., 2012; Tabib-Salazar et al., 2013).

The structural architecture of isolated RbpA has been defined by secondary structure predictions and solution NMR (Bortoluzzi et al., 2013; Tabib-Salazar et al., 2013) (**Figure 1.8**). A central core domain (CD) comprising a β-barrel fold is flanked by an unstructured 26 amino acid N-terminal tail and a C-terminal segment predicted to harbor two α-helices linked to the CD by a 15-residue basic linker (BL). RbpA has been shown to directly interact with σ_2 of group 1 and group 2 σ-factors. This interaction is mediated by the protein's C-terminal domain (which we designate the Sigma Interaction Domain, SID), and point mutations that disrupt σ-binding also disrupt RbpA function (Tabib-Salazar et al., 2013).

RbpA has been shown to activate transcription from a wide range of *Sco* σ^{HrdB}-, *Sco* σ^{HrdA}-, *Mtb* σ^{A}-, and *Mtb* σ^{B}-dependent promoters *in vitro* (Hu et al., 2014; 2012; Tabib-Salazar et al., 2013) and, like CarD, co-localizes with initiation complexes *in vivo* (Tabib-Salazar et al., 2013); However, in comparison to CarD, much less is known about the RbpA mechanism for transcription activation.

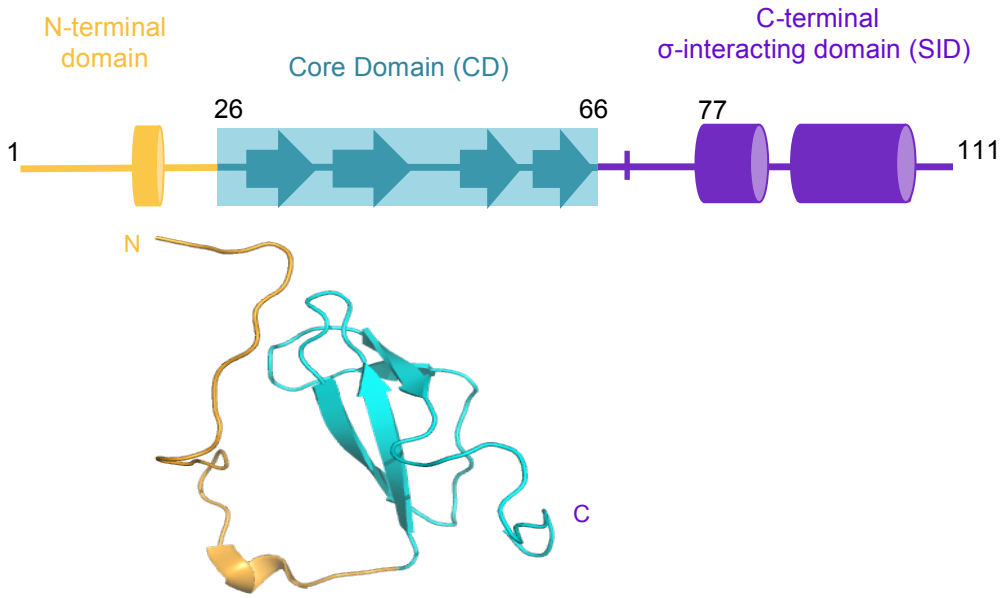


Figure 1.8: Structural architecture of mycobacterial RbpA. Top: RbpA contains a central core domain (CD) comprising a β-barrel fold that is flanked by an unstructured 26 amino acid N-terminal tail and a C-terminal segment predicted to harbor two a-helices linked to the CD by a 15 residue basic linker (BL). Bottom: NMR structure of CD and NTD (Tabib-Salazar et al., 2013).

1.9 Prokaryotic RNAP as a drug target

RNAP is the target for a number of small molecule inhibitors. The rifamycin class of antibiotics is one the most potent and broad-spectrum antibiotic and is a first-line therapeutic treatment for tuberculosis. A crystal structure of the antibiotic rifampicin (Rif) bound to *Taq* RNAP revealed the structural mechanism for Rif inhibition. Rifampicin binds a pocket of the β subunit within the RNAP main channel, about 12 Å away from the active site, and directly blocks the path of the elongating RNA transcript. As a result, the RNAP is only able to produce short transcripts 2-3 nt in length (Campbell et al., 2001; Feklistov et al., 2008).

In addition to Rifamycins, many other small molecule antibiotics have been identified that inhibit RNAP, targeting various regions of the enzyme, including lipiarmycin (fidaxomicin), which is effectively used for the treatment of *Clostridium. dificile* infection. The RNAP is a large molecular machine with a complex functional cycle, and it is likely that there are additional targets on the enzyme that have not been exploited. That RNAP is an attractive and proven drug target for anti-tuberculosis therapies highlights the importance of having both a structural and functional understanding of mycobacteria RNAP.

Chapter 2:

Structural and Functional Analysis of the RbpA SID and BL

RbpA, an Actinobacterial-specific transcription activator that is essential in *Mtb* (Forti et al., 2011) and necessary for normal growth in *Sco* (Newell et al., 2006), has been shown to co-localize at initiation complexes with the primary σ factor (σ^{HrdB}) in *Sco in vivo* (Tabib-Salazar et al., 2013) and to stimulate transcription from a range of *Mtb* σ^{A} (group 1) and σ^{B} (group 2)—dependent promoters *in vitro* (Hu et al., 2014; 2012; Tabib-Salazar et al., 2013). Interactions between RbpA and core RNAP have been proposed previously, and putative interactions have been mapped to two widely spaced regions of RNAP (Dey et al., 2011; Hu et al., 2012). However, in the absence of core RNAP, RbpA forms a stable binary complex with the σ_2 domain of housekeeping σ-factors and has been shown to interact with group 2 σ-factors, but not group 3 or group 4 σ factors (Hu et al., 2012; Tabib-Salazar et al., 2013).

NMR structures comprising the RbpA flexible NTD and structured CD composed of β-sheets (described in the introduction) have been reported (Bortoluzzi et al., 2013; Tabib-Salazar et al., 2013); however, these structures did not resolve the C-terminal SID that is responsible for binding σ, and thus provide few clues to RbpA's function. In this chapter, I report the 2.2Å resolution crystal structure of the *Mtb* RbpA/σ^{A}_{2} complex. The structure reveals the molecular basis of the RbpA-SID interaction with σ^{A}_{2} and the basis of RbpA binding specificity for group 1 and group 2 σ factors. Further, the structure allowed us to generate a model of the RbpA-SID in the context of a transcription initiation complex, which suggests that RbpA contacts promoter DNA. Using a

fluorescence-based binding assay, I tested and confirmed that RbpA plays a role in modulating RNAP-DNA binding.

2.1 The importance of the SID in transcription activation

Using bacterial-2-hybrid experiments, Mark Paget's group previously determined that the RbpA's C-terminal domain (SID) was responsible for binding σ, identifying the importance of this domain for RbpA function (Tabib-Salazar et al., 2013). In collaboration with the Darst lab, they found that an RbpA truncation lacking the SID and the BL (a 15-residue basic linker between the SID and the CD) was unable to activate transcription *in vitro* (Hubin et al., 2015). They also found that a SUMO-tagged RbpA construct comprising just the SID-BL (SUMO-RbpA^{SID-BL}) partially activated transcription, but that full length RbpA was necessary for full activation (Hubin et al., 2015). In independent experiments, however, I investigated an untagged RbpA comprising just the SID-BL (RbpA^{SID-BL}) and found that it activated as well as, if not better than, WT RbpA in *in vitro* transcription assays (data presented in Chapter 4). In this chapter I describe the structure of the SID-BL bound to σ^A, while further biochemical and kinetic analysis of RbpA's domain function will be discussed in Chapters 4 and 5.

2.2 Crystallization of RbpA-σ^A_2 complex

For crystallization of an RbpA-σ_2 complex, I expressed His_6-SUMO-σ^A_2 fusion (σ residues 224-364) and RbpA from a co-expression cassette generated by Mark Paget's group (**Figure 2.1a**). The proteins formed a tight complex that could be co-purified by nickel and gel filtration columns (**Figure 2.1b**). The His_6-SUMO tag was designed to be

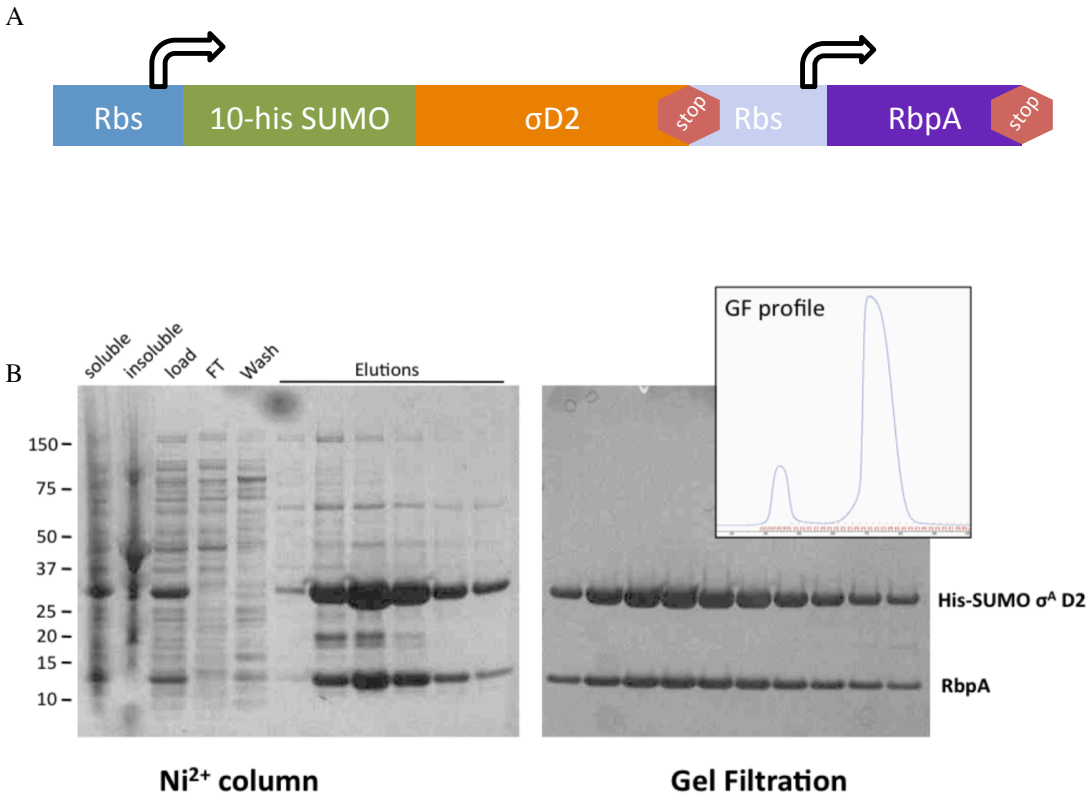


Figure 2.1: Expression and purification of the complex between *Mtb* RbpA and σ^{A}_{2}. (A) Recombinant proteins were overexpressed in *E. coli* cells from a co-expression plasmid containing σ^{A}_{2} with a cleavable his-SUMO tag and untagged RbpA. (B) A complex between RbpA and his-SUMO tagged σ^{A}_{2} was purified through nickel column and gel filtration (SD200).

cleavable from σ^{A}_{2} by ULP1 protease; however, upon removal of the tag, the complex precipitated. I therefore kept the tag on σ^{A}_{2} throughout purification and crystallization trials.

The complex was screened for crystallization conditions, and diffracting crystals of similar morphology appeared under a number of different conditions. SDS-PAGE gels of these crystals showed that proteolysis occurred within the drops, and that they contained two major protein fragments (**Figure 2.2a**). MALDI-TOF analysis of the crystal contents revealed that both proteins were N-terminally proteolytically degraded, resulting in crystals containing a mixture of multiple σ^{A}_{2} fragments (lacking up to 18 N-terminal residues) and multiple RbpA fragments (all including the BL and SID) (**Figure 2.2b-c**).

A 2.2 Å data set was collected on crystals from one of the conditions (0.1 M Tris pH 8.5, 0.5 M ammonium sulfate (**Figure 2.3a-b**). The data were phased by molecular replacement with an *Mtb* σ^{A}_{2} homology model (ExPASy SWISS MODEL) based on a previously solved structure of the corresponding domain from *Taq* σ^{A} (1KU2, Campbell et al., 2002) (**Figure 2.3c**). Electron density maps revealed the RbpA-SID (*Mtb* residues77-108) bound to σ^{A}_{2} (*Mtb* residues 242-363) (**Table 2.1**). Although the MALDI-TOF analysis indicated that the mixture of RbpA crystallized fragments included some intact RbpA molecules as well as RbpA molecules lacking only the NTD, additional electron density was absent, and the rest of RbpA was presumed disordered or absent.

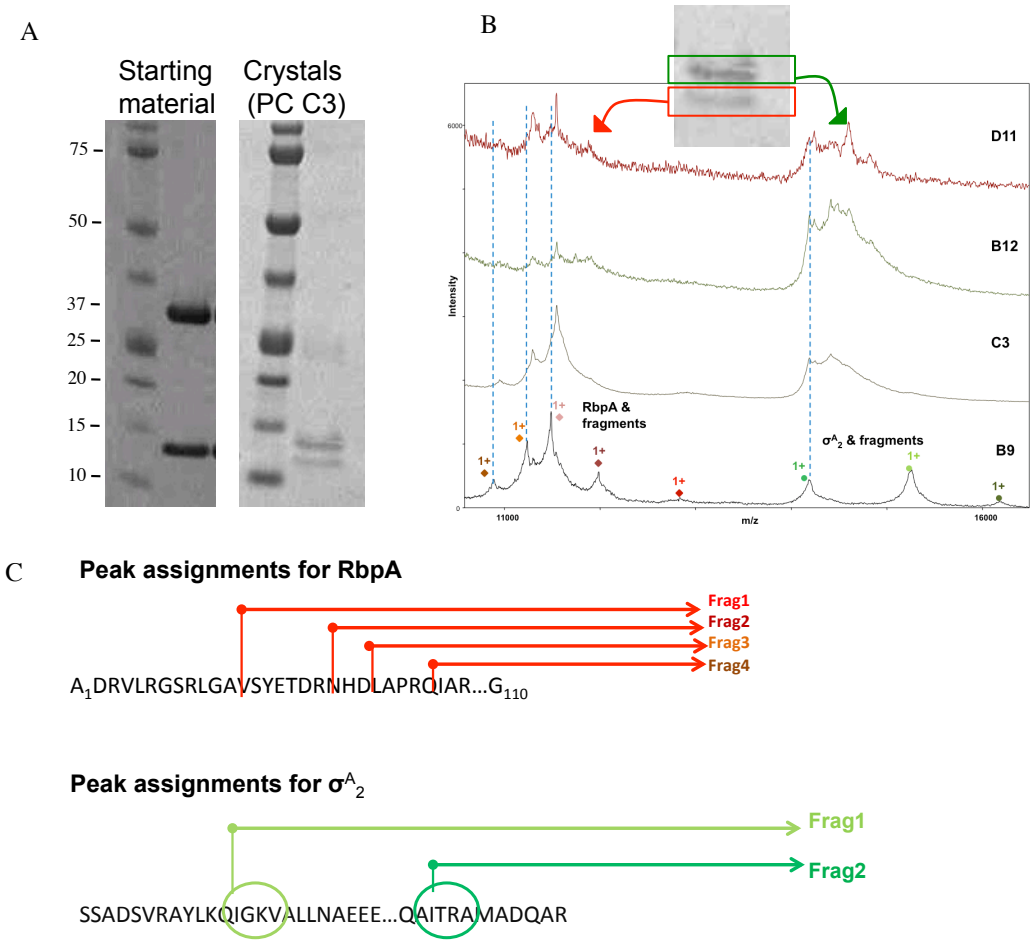
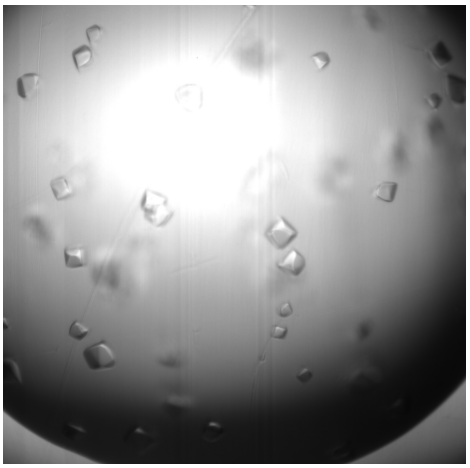
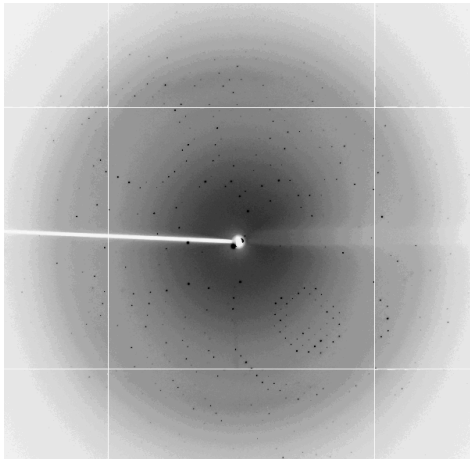


Figure 2.2: Mass spectrometry indicates crystals are proteolytically trimmed RbpA-σ^A_2 complex. (A) Gel of crystals shows two protein fragments that did not match molecular weight of the his_6-SUMO-σ^A_2/RbpA complex (B) matrix-assisted laser desorption ionization (MALDI) mass spectrometry profiles of four crystals (D11, B12, C3, B9). Mass groups around 11.5 kDa and 14.5 kDa match gel bands, and correspond to RbpA and σ^A_2 fragments, respectively. (C) Sequence assignment of fragments within the crystals. Mass heterogeneities indicate both proteins are undergoing varying degrees of protein degradation at their N-terminus.

A



B



C

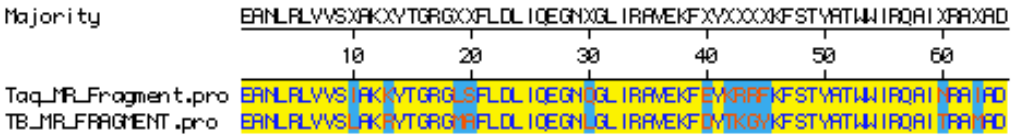


Figure 2.3: Crystals and data collection of σ^{A}_{2}/RbpA complex. (A) Image of crystals diffracted. (B) Diffraction pattern. Data was collected at BNL NSLS x-29 beam where crystals diffracted to 2.2 Å. (C) Alignment between *Taq* and *Mtb* σ^{A}_{2}. Data were phased with a homology model made using the structure of *Taq* σ^{A}_{2}.

Table 2.1: Crystallographic Statistics for σ^{A}_{2}/RbpA complex

Data collection	
Space group	$P3_121$
Combined datasets	1
Cell dimensions	
a, b, c	84.459, 84.459, 73.701
α, β, γ	90, 90, 120
Wavelength (Å)	1.74
Resolution (Å)	36.64–2.202 (2.282–2.202)[a]
Total reflections	182,175 (9,287)
Unique reflections	15,679 (1,423)
Multiplicity	11.6(6.2)
Completeness (%)	99 (96)
$\langle I \rangle / \sigma I$	22.25 (2.07)
Wilson B-factor	42.40
R_{merge}[b]	0.1189 (0.7155)
R_{meas}[b]	0.1244 (0.7709)
CC1/2[b]	0.997 (0.998)
CC*[b]	0.999 (0.999)
Refinement	
R_{work} / R_{free}	0.1951/0.2310 (0.2864/0.2966)
CC_{work}/CC_{free}[c]	0.941/0.919 (0.818/0.841)
No. non-hydrogen atoms	
Protein	1260
Ligand/ion	19
Water	58
Protein residues	154
B-factors	
Protein	56.41
Ligand/ion	96.27
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.7
Clashscore	5.88
Ramachandran favored (%)	98
Ramachandran outliers (%)	0

CC, correlation coefficient.
Values in parentheses are highest resolution shells

The crystal structure reveals that the RbpA-SID comprises two α-helices (α1, α2; **Figure 2.4**), confirming previous sequence-based structural predictions of the domain (Tabib-Salazar et al., 2013). In addition to the helical SID, four residues of the 15-residue RbpA-BL connecting the RbpA-SID to the RCD are also visible in the structure. Both of the RbpA-SID α-helices contact $\sigma^A{}_2$, forming a significant intermolecular interface with a buried surface area of 948 $Å^2$.

The RbpA-SID makes extensive contacts with residues from both the conserved regions of $\sigma^A{}_2$ (regions 1.2 and 2.3) and the NCR (**Figure 2.5a-b**). Two conserved arginine residues in the RbpA-SID (*Sco* RbpA R89 and R90, corresponding to *Mtb* R88 and R89) that are critical for σ binding were identified in previous studies (Tabib-Salazar et al., 2013). The structure shows that these two residues form extensive electrostatic interactions with σ^A (**Figure 2.5c**), explaining the mutagenesis studies. Specifically, *Mtb* RbpA-R88, located in α1 of the SID, forms a salt bridge with σ^A residue E254, while RbpA-R89, located in the short linker between the two SID α-helices, makes salt bridges with both E254 and D336 of σ^A.

A cluster of conserved hydrophobic RbpA residues (L94, L97, and L98; Figs. 2B, C), located on RbpA-SID α2, make extensive van der Waals contacts with residues of σ^A (**Figure 2.5c-d**). The results of bacterial two-hybrid (BTH) assays with a *Sco* RbpA mutant in which the residues corresponding to *Mtb* RbpA L97 and L98 (*Sco* RbpA V98 and L99) were changed to alanine confirmed the importance of interactions observed in the crystal structure, revealing that these branched hydrophobic residues are necessary for *Sco* RbpA/σ^A binding.

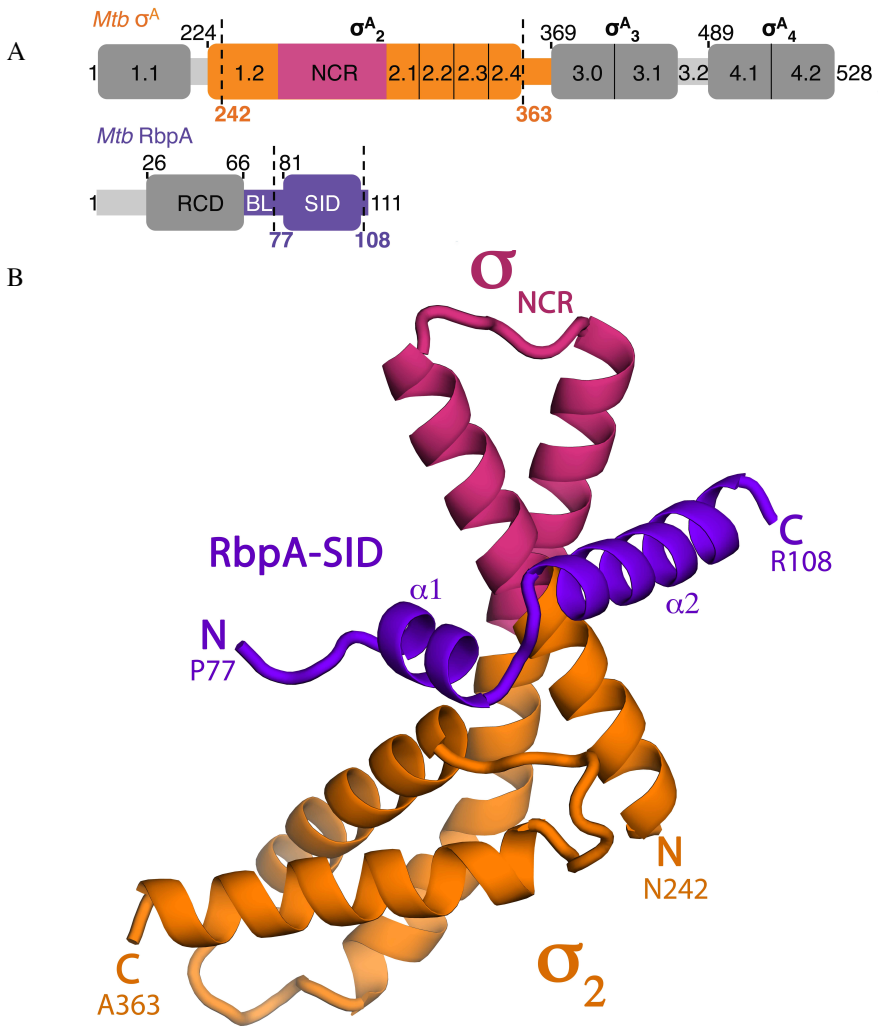
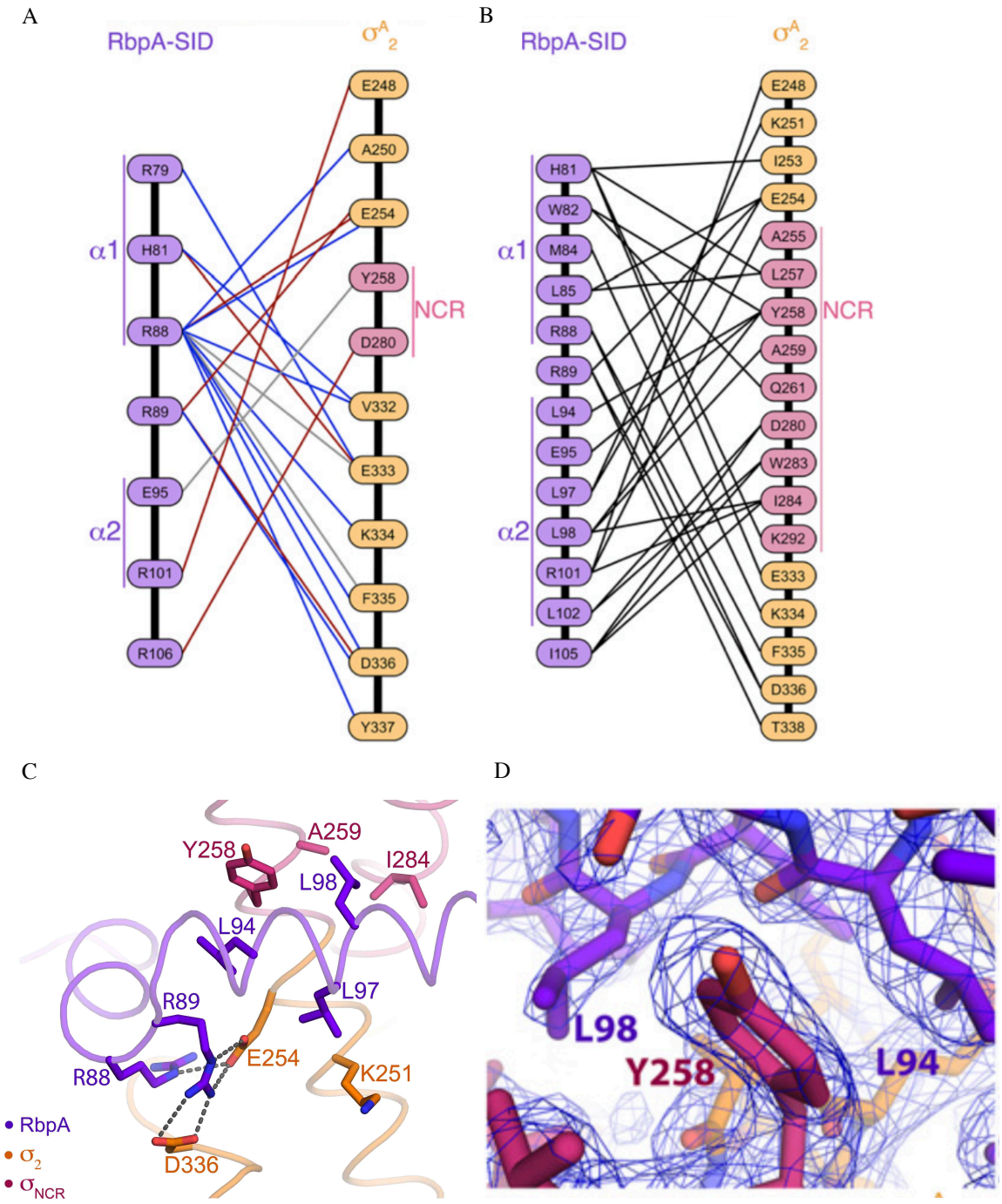


Figure 2.4: Structure of RbpA-SID-σ^A_2. (A) Schematic of RbpA and σ^A architecture. Domain 2 of σ^A is shown in orange with the NCR in fuschia and the remaining regions colored gray. The SID and BL of RbpA are colored purple and the remaining regions are gray. Regions visible in the crystal structure are flanked by dotted lines. (B) Crystal structure of RbpA-SID-σ^A_2 is shown in ribbon and colored as in (A). Adapted from Hubin et al., 2015.

Figure 2.5 (Following page): RbpA contacts residues from both the conserved region and NCR of σ^{A}. (A) Schematic highlighting polar and ionic molecular interactions between the RbpA SID and σ^{A}_{2}. Interactions between residues are indicated by lines. Ionic interactions are indicated red, hydrogen bonds are gray, and hydrogen bonds mediated through water are blue. (B) Schematic showing nonpolar interactions between RbpA SID and σ^{A}_{2}. (C) Highlighted RbpA-SID residues required for σ-binding. (D) Selected portion of a simulated annealing composite omit electron density map at 1σ. Clear density for RbpA is shown. RbpA (purple) and σ^{A}_{2} (orange with the NCR in fuschia) are shown in stick, highlighting the hydrophobic interactions between RbpA residues L94 and L98 with σ^{A}_{NCR}-Y258. Adapted from Hubin et al., 2015.


A
RbpA-SID
σA2
α1
α2
NCR
R79
H81
R88
R89
E95
R101
R106
E248
A250
E254
Y258
D280
V332
E333
K334
F335
D336
Y337
B
RbpA-SID
σA2
α1
α2
NCR
H81
W82
M84
L85
R88
R89
L94
E95
L97
L98
R101
L102
I105
E248
K251
I253
E254
A255
L257
Y258
A259
Q261
D280
W283
I284
K292
E333
K334
F335
D336
T338
C
A259
Y258
I284
L98
L94
L97
R89
E254
R88
K251
D336
RbpA
σ2
σNCR
D
L98
Y258
L94

The RbpA-SID/σ^A_2 structure provides a basis to understand why RbpA selectively binds to group 1 and 2 σ factors but not group 3 σ factors or EFC σ factors. An alignment of σ_2 amino acid sequences of *Mtb* σ^A (group 1) and σ^B (group 2) with *Mtb* σ^F (ECF) (**Figure 2.6**) revealed that out of 21 *Mtb* σ^A residues contacting RbpA, 16 were identical in *Mtb* σ^B while only four were identical in *Mtb* σ^F, explaining RbpA's σ selectivity.

2.3 Model of RbpA on holoenzyme reveals DNA binding role

While the RbpA-SID/σ_2 structure provided specific details of RbpA's interaction with σ, I obtained further insight into the role of RbpA in transcription initiation by modeling the factor on an initiation complex. We generated a structural model of RbpA bound to RPo by superimposing conserved regions of σ^A_2 from the *Mtb* RbpA-SID/σ^A_2 complex onto the corresponding regions of *Taq* σ^A_2 in an RPo model (Feklistov and Darst, 2011; Murakami et al., 2002) (0.681 Å root-mean-square deviation over 93 Cα atoms) (**Figure 2.7a**), resulting in an RbpA/RPo model with no steric clashes.

In the model, the interaction between the RbpA-SID and σ^A_2 positions the RbpA-BL on the minor-groove side of the duplex promoter DNA just upstream of the -10 element. Conserved RbpA-R79 (**Figure 2.7c**) is positioned to make electrostatic interactions with the non-template strand (nt-strand) DNA phosphate backbone at the -14 position with respect to the transcription start site at +1 (**Figure 2.7b**). In addition, conserved RbpA-M84 is positioned to potentially play a role in DNA binding through van der Waals interactions. Also of note is that in the RbpA-BL, just N-terminal to the modeled portion of RbpA, are three lysines residues (K73, K74, K76, **Figure 2.7c**), which may play a role in forming electrostatic interactions with the DNA backbone.

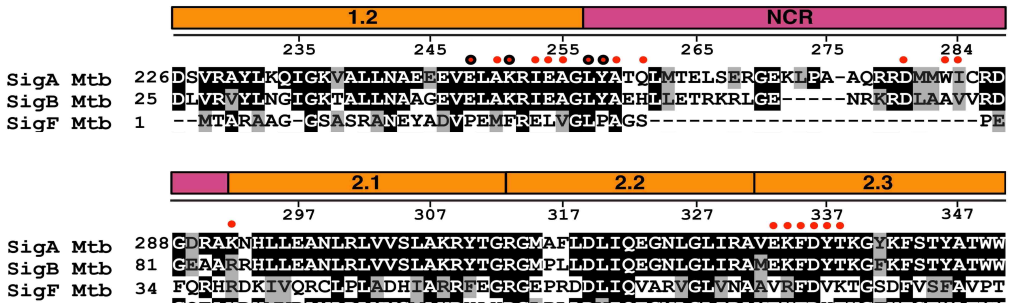


Figure 2.6: Contacts between RbpA and σ^A reveals basis for σ specificity of RbpA. Sequence alignment of regions 1.2-2.3 in *Mtb* σ^A, $\sigma^B$and σ^F. Amino acid groups were: VLMA, RK, ED, YFW, TS and QN. Remaining residues (H, C, P, and G) were not grouped. σ^A residues that contact σ^A are indicated by red dots. Adapted from Hubin et al. 2015.

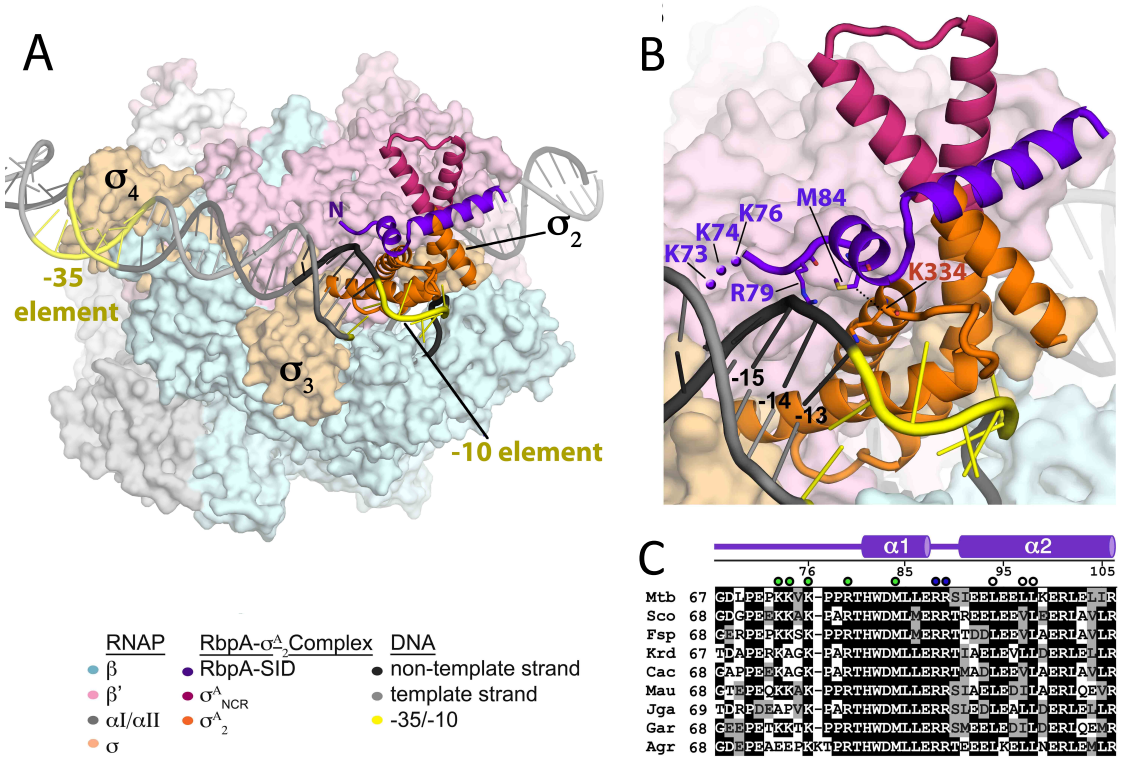


Figure 2.7: Structural model of the RbpA-SID/RPo suggests DNA binding role of RbpA. (A) Structural model of RbpA bound to RPo was generated by superimposing conserved regions of *Mtb* σ^A_2 (from the *Mtb* $RbpA_{SID}/\sigma^A_2$ structure) with *Taq* σ^A in an RPo model. Protein and DNA elements are colored as indicated. RNAP is shown as a molecular surface, RbpA-SID/σ^A_2 is shown in ribbon, and DNA is shown as a phosphate backbone worm. (B) Magnified view of model shows RbpA is positioned to contact DNA upstream of promoter element. RbpA amino acids R79, located in the BL, and M84, located in α1, are in position to contact the nt-strand phosphate backbone at the -13/-14 position. (C) RbpA sequence reveals three conserved, positively charged lysine residues K76, K74, K73 (represented by purple dots) located in the RCD-SID linker that would be well positioned to interact with the negatively charged DNA phosphate backbone (depicted as dots in (B)). Adapted from Hubin et al. 2015.

The hypothesis of an RbpA/DNA interaction was validated using formaldehyde cross-linking, which links atoms just 2 Å apart (**Figure 2.8)** (Hubin et al., 2015)**.**

When RbpA was added to an *Mbo* holoenzyme-fork DNA complex, an additional crosslink appeared, which was confirmed to be an RbpA-DNA crosslink since a slower migrating band appeared when the assay was repeated with SUMO-RbpA. RbpA-DNA cross-links were also detected with SUMO-RbpA(72-111), which includes the RbpA-BL and SID but not with SUMO-RbpA(1-71), which lack the RbpA-SID and BL essential for σ^A binding and activity.

2.4 RbpA increases holoenzyme affinity for promoter DNA

An RbpA/promoter DNA interaction would be predicted to increase the overall affinity of RNAP for promoter DNA. To test this hypothesis, I designed a fluorescence anisotropy assay to measure RNAP binding to promoter DNA with and without RbpA (Owen and McMurray, 2009). When a fluorophore is unbound or bound to a small molecule, it will tumble in space rapidly and emit largely depolarized light. However, when the fluorophore is bound to a larger complex, the rate of tumbling decreases, and the light emitted by the fluorophore becomes increasingly polarized. Thus, using a cy3-labeled promoter DNA, I was able to quantify RNAP binding to promoter DNA and determine binding constants both with and without RbpA.

Using the fluorescence anisotropy assay, I measured the binding of RNAP to a Cy3-labeled duplex VapB promoter template (**Figure 2.9a**). I generated binding curves by performing the assay at varying RNAP concentrations while keeping the DNA concentration constant (**Figure 2.9b**). I found that the addition of RbpA increased the

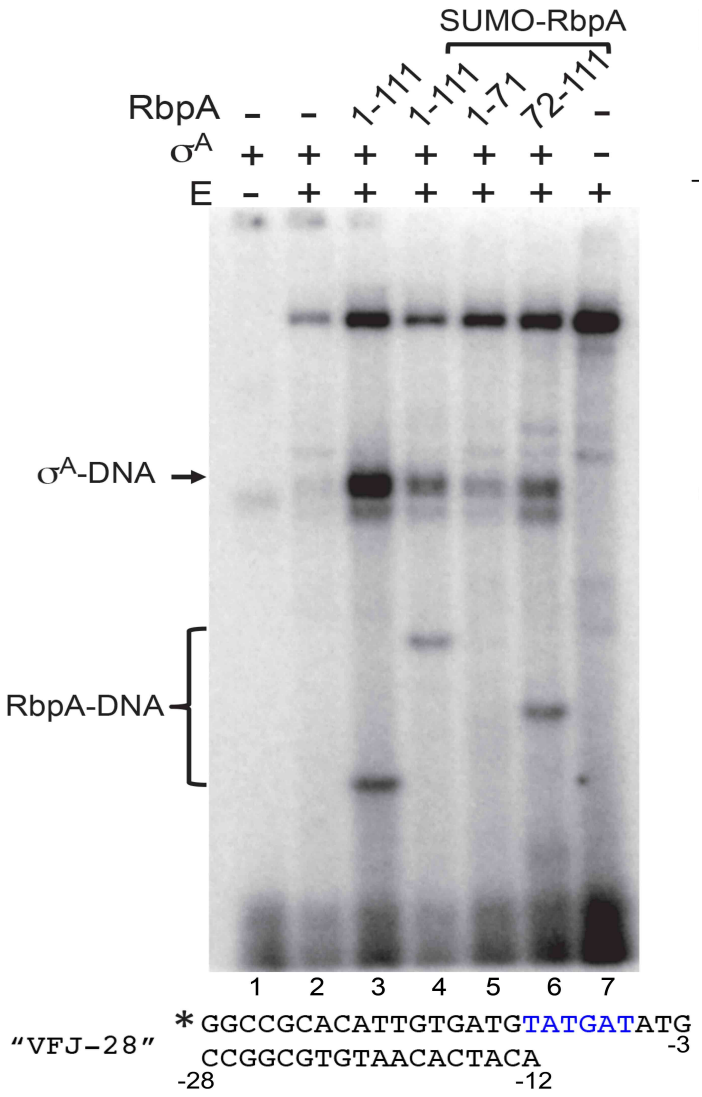


Figure 2.8: RbpA cross-links to fork junction promoter DNA. Reactions contained VFJ-28 DNA radiolabeled at the 5' end of the non-template strand, core *Mbo* RNAP (E) at 200 nM, σ^A (1 μM), and native or SUMO-fused derivatives of RbpA (2 μM), as indicated. Complexes were allowed to form at 37°C for 15 min before treatment with formaldehyde. Species were separated by SDS-PAGE and visualized by phosphorimaging. Below is the *vapB10* promoter-based fork junction DNA template VFJ-28, indicating the -10 element (blue) and labeled non-template strand (asterisk). Experiment was performed in collaboration with Mark Paget's group at University of Sussex. Adapted from Hubin et al. 2015.

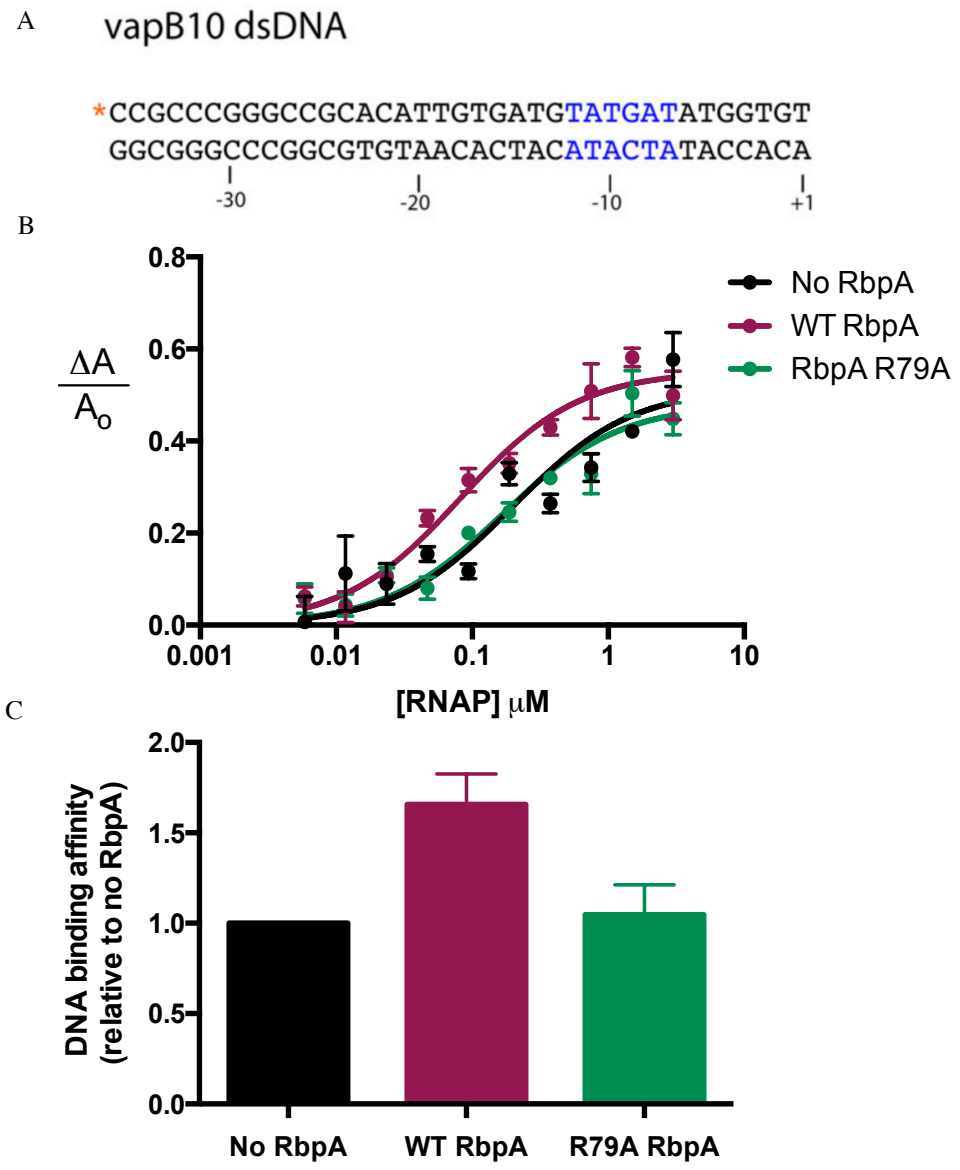


Figure 2.9: RbpA increases RNAP affinity for promoter DNA. (A) Cy3-labeled VapB dsDNA used in anisotropy assay. (B) Representative fluorescence anisotropy binding curves showing RNAP binding to Cy3-labeled VapB promoter DNA (200 nM). (C) WT RbpA increases affinity of RNAP for promoter DNA while RbpA-R79A has little-to-no effect on binding. Error bars are based on average fold change from nine experiments. Adapted from Hubin et al., 2015.

affinity of RNAP for promoter DNA (indicated by a decrease in the dissociation constant, K_D) nearly two-fold (**Figure. 2.9b-c**), consistent with the modest activation activity of RbpA in abortive initiation assays on the VapB promoter (Hubin et al., 2015). On the other hand, an RbpA-R79A mutant had no significant effect on RNAP binding to promoter DNA (**Figure 2.9b-c**), providing further evidence that RbpA-R79 plays an important role in DNA binding. This mutant purifies normally on size exclusion columns and co-purifies with σ, indicating that the lack of an effect of the R79A mutant on DNA binding is not because of improper protein folding.

Upon DNA binding, a loop in region 2.3 of σ flips out the A_{-11} base in the -10 element. The *Mtb* RbpA/σ^A_2 complex indicates that RbpA makes several direct contacts to residues of the 2.3 loop, and the loop conforms to the DNA-binding conformation, suggesting RbpA may allosterically stabilize σ in a DNA binding conformation. I tested if this were the case by measuring RbpA's effect on the holoenzyme binding short ssDNA comprising just the -10 element. RbpA had no significant effect on RNAP binding to Cy3-labeled single-stranded DNA comprising only the -10 and discriminator elements (**Table 2.2**), supporting the hypothesis that the effect of RbpA on RNAP-promoter binding is through interactions of RbpA with duplex DNA upstream of the -10 element.

To test whether the mechanism for transcription activation by RbpA is based on RbpA-DNA contacts upstream of the -10 element, our collaborators in Mark Paget's group tested the R79A in *in vitro* run-off transcription assays and found that it did not stimulate transcription (Hubin et al., 2015). Similarly, I found that the R79A completely failed to activate in abortive initiation assays (**Figure 2.10**). The mutant completely failed

Table 2.2 Average K_D values of RNAP binding to Cy3-labeled ssDNA (-12 to +1) with and without RbpA

Promoter	Sequence	No RbpA K_D (μM)	RbpA K_D (μM)
Consensus	*TATAATGGGAAGG	0.38 ± 0.07	0.33 ± 0.04
AP3	*TAGACTGGCAGGG	2.24 ± 0.41	1.88 ± 0.31
vapB	*TATGATATGGTGT	1.62 ± 0.26	1.77 ± 0.15

Average K_Ds of RNAP binding to Cy3 labeled ssDNA (-12 to +1) were determined using fluorescence anisotropy. The addition of RbpA to the assay has no significant effect on RNAP binding to ssDNA composed of the -10 and discriminator elements. -10 element is highlighted in blue. Error is based off of triplicate trials.

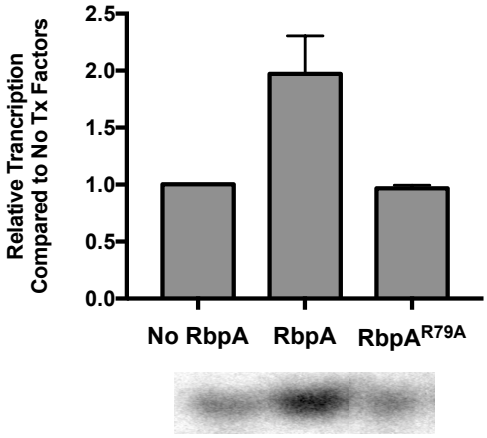


Figure 2.10: RbpA-R79 is a critical residue for RbpA activation. An RbpA mutant, $RbpA^{R79A}$ fails to activate transcription on VapB in abortive initiation transcription assays, indicating this residue is essential for RbpA's function.

to stimulate transcription from a VapB promoter template, indicating that RbpA's role in DNA binding is essential for its activation function and, more specifically, R79 is critical for RbpA's mechanism of activation. Further analysis of the R79A mutant will be discussed in Chapters 4 and 5.

2.5 Discussion

The RbpA-SID/σ^{A}_{2} structure provides clues into RbpA's function and activation mechanism. The interaction between the SID and σ determines the specificity of RbpA for group 1 and group 2 σ factors. Our structural modeling indicates that the SID likely plays a key role in positioning conserved BL residue R79 near the upstream edge of the transcription bubble. The favorable protein/DNA contacts added to the transcription initiation complex would potentially stabilize a transcription initiation intermediate, which is consistent with findings that RbpA activates transcription primarily by stimulating RPo formation. Indeed, fluorescence anisotropy revealed RbpA increases RNAP affinity for DNA, and R79 is key for its function.

The complex represents the first structure of a transcription factor interacting with the NCR of a housekeeping σ. Interacting with the primary σ-NCR is a property that RbpA shares with the unrelated *Chlamydia trachomatis* (*Ctr*) transcription factor GrgA, which binds to the *Ctr* σ^{66}-NCR and DNA (Bao et al., 2012). The holoenzyme assembly factor Crl from enteric bacteria, although structurally distinct from RbpA, interacts with σ_2 of the group 2 sigma factor σ^{S} (Banta et al., 2013; 2014).

While previous studies have suggested the NCR modulates promoter escape and pausing (Leibman et al., 2007), the role of the NCR region has remained, for the most

part, unknown. The discovery of factors that bind the NCR, such as RbpA, has shed light on its role in initiation and gene regulation. RbpA interaction with residues in both the conserved and the NCR determines its specificity for group 1 and certain group 2 σ factors. Based on our structural modeling, it's likely that the RbpA NTD, CD, or both form additional interactions with RNAP. However, our RbpA-SID/RPo structural model is completely incompatible with previous studies claiming that RbpA interacts with RNAP either in the active-site channel near the Rif binding site (Dey et al., 2010; **Figure 2.11a**) or with a different region on the β-subunit (Hu et al., 2012; **Figure 2.11b**). Based on the location and orientation of the N-terminus of the RbpA-SID in our structural model, it would be impossible for the regions of RbpA that were identified to bind to core RNAP to span the distance to either of the putative binding sites. The binding determinant near the Rif-binding pocket was originally inferred based on functional data and then refined based on a single crosslink (Dey et al., 2010).

The other β-subunit binding region was identified using cleavage experiments through hydroxyl-radicals generated from Fe-BABE attached to the lone, internal Cys-residue of the RbpA-RCD (Hu et al., 2012). It's plausible that mutation of the internal Cys residue, buried in the CD of RbpA, caused misfolding or aggregation of RbpA, resulting in Fe-BABE cleavage reflecting non-specific interactions with the RNAP surface. Our RbpA-SID/RPo model suggests that the likely RNAP binding determinant is on the β' subunit on the top of the clamp domain.

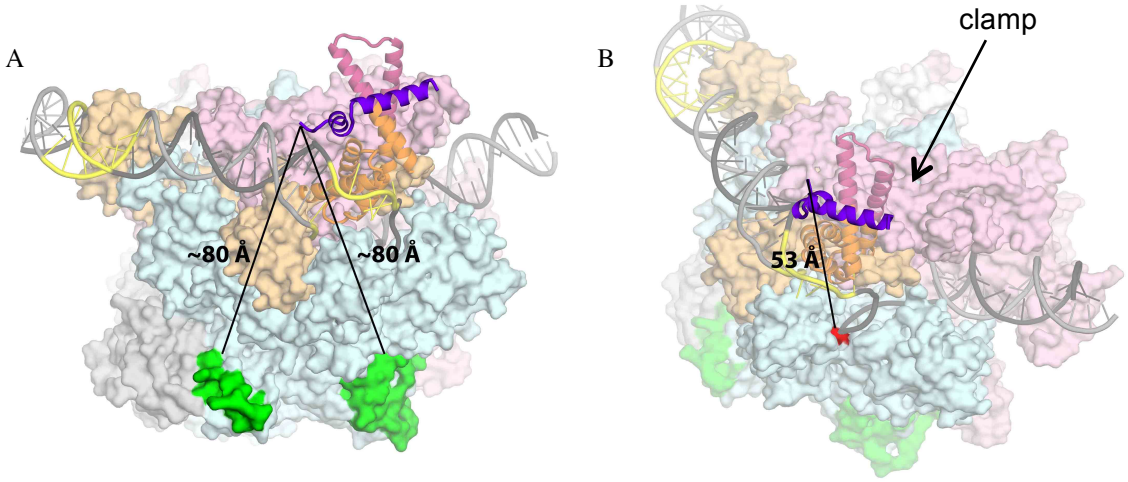


Figure 2.11: Proposed RbpA binding sites on RNAP core are incompatible with current structure. Protein and DNA elements are colored as in Figure 2.7 (A) The proposed RbpA interacting residue R381 (red) on RNAP β subunit was predicted through crosslinking to bind K74 in the BL (Dey et al., 2010). However, RbpA-P77, just 4 residues away from K74, is 53 Å away from βR381 (direct Cα-Cα distance). (B) The RbpA-RCD binding regions predicted by iron-babe cleavage (green) (Hu et al., 2012) are located at least 80Å away from the beginning of the RCD-SID linker (measured as a straight line from the Cα of P77).

Chapter 2 Acknowledgements:

Crystallization trials were performed with the assistance of Joshua Flack, a summer research student in the Darst Lab. Certain transcription assays (specified in text), bacterial-2-hybrid experiments, and crosslinking experiments were performed by Mark Paget's group, and the MALDI was performed by Paul Dominique B. Olinares in the Chait Laboratory at The Rockefeller University.

Chapter 3:

Structural Studies of Mycobacterial RNAP-RbpA Complexes

Although the RbpA-SID/σ_2 structure provided clues into RbpA function, a more complete understanding of the RbpA mechanism requires a crystal structure of full-length RbpA bound to its target, a transcription initiation complex. In this chapter, I summarize my crystallization experiments with *Mbo* and *Msm* RNAPs and describe the 2.76Å resolution crystal structure of the *Msm* RNAP holoenzyme bound to promoter DNA and in complex with full-length RbpA (*Msm* holo-RbpA-DNA). The structure reveals that the RbpA-CD binds to the regions of β' defined as the zipper and zinc-binding domain (ZBD). It also shows that the BL contacts promoter DNA upstream of the transcription bubble, confirming our structural model described in Chapter 2.

To date, all published bacterial RNAP crystal structures have been from *Eco* or *Thermus* species. Therefore, the *Msm* holo-RbpA-DNA structure not only offers further insight into RbpA's function, but it also provides a structural understanding of mycobacterial RNAP as well as invaluable information for future biochemical studies investigating regulation of the mycobacterial enzyme. Further, the structure is the highest reported resolution structure of a bacterial RNAP-promoter initiation complex, affording greater insight into key molecular interactions involved in transcription initiation. An analysis comparing *Taq* and *Eco* initiation complexes will be discussed in Chapter 6, and future use of the *Msm* holo-RbpA-DNA crystals as a platform for studying anti-tuberculosis therapies will be discussed in Chapter 7.

3.1 Crystallization of Mbo holo-RNAP

To obtain a stoichiometric *Mbo* Holo-RbpA protein complex for crystallization, we first co-expressed *Mbo* σ^A and *Mtb* RbpA (100% identical to *Mbo* RNAP) and purified the σ^A/RbpA complex. We purified *Mbo* RNAP using methods described by Davis et al., 2015; however, prior to the final gel filtration column, σ^A/RbpA was added in excess, resulting in a stoichiometric complex after gel filtration (**Figure 3.1a**). The complex was screened with multiple upstream fork DNA constructs (**Figure 3.2)**, and promising micro crystals formed in a number of conditions. We were able to reproduce and optimize crystals and microcrystals from one condition containing an *Mtb* holo-RbpA-T6 fork complex (**Figure 3.1b**). The optimized crystals (grown in 0.2 M ammonium sulfate, bis-tris pH 6.5, 20% PEG 3350) were thin plates that diffracted to ~8Å (**Figure 3.1c**). Data was collected on these crystals, but difficulty indexing and incomplete data caused issues with data processing. Further optimization of these crystals is a priority for future studies.

3.2 Purification and Crystallization of *Msm* holo-RbpA:

For more than a decade, the Darst lab has used endogenous protein preps of *Taq* RNAP to yield high quality protein that produces well-diffracting crystals. Endogenous protein purification isolates proteins from their native bacterial cells, and unlike recombinant protein preps, does not rely on over-expression of protein in *E. coli,* which can cause aggregation, misfolding, and degradation. Because a large amount of cell mass is required for endogenous purification, this method would not be feasible for purifying

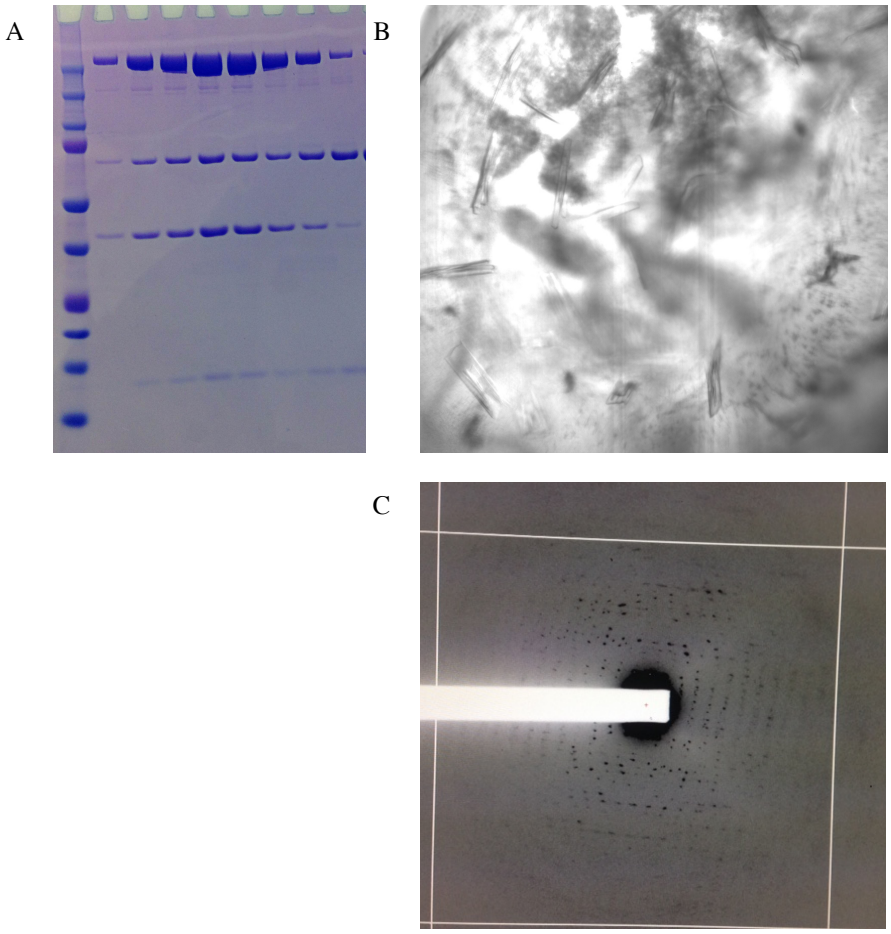


Figure 3.1: Crystallization of *Mbo* holo-RbpA with T6 fork DNA. (A) SDS-PAGE gel of purified *Mbo* holo-RbpA complex (B) Crystals after initial optimization (C) Diffraction of crystals at APS NE-CAT beam source. Crystals diffracted to ~8Å.

Figure 3.2: Fork DNA constructs used in crystallography trials

RNAP from *Mtb* due to the slow growth of the organisms and its classification as biosafety level 3 (BSL 3) by the Center for Disease Control. We therefore decided to endogenously purify RNAP from *Msm*, which is highly conserved to that of *Mtb*.

I developed and optimized a purification protocol of an endogenous *Msm* RNAP from an *Msm* strain containing a his_{10}-tag fused to the native chromosomal *rpoC* gene (*rpoC*::*rpoC*-his10 *Msm)* produced by Allison Faye in Mike Glickman's lab. The *rpoC*::*rpoC*-his10 *Msm* cells were grown and pelleted by the Bioexpression & Fermentation Facility, University of Georgia. After polyethyleneimine (PEI) and ammonium sulfate precipitation followed by nickel and biorex columns, the protocol yielded a mixture of holo and core RNAP with a contaminating protein about 30 KD in size (**Figure 3.3;** Chapter 8). N-terminal sequencing of this band (performed by University of Texas Medical Branch facilities) revealed that it is the ECF σ-factor σ^F. I added a 5-fold molar excess of purified *Msm* σ^A-RbpA to the *Msm* RNAP sample prior to a final gel filtration column, with the goal of displacing the contaminating σ^F with the desired σ^A-RbpA. This resulted in a highly pure, stoichiometric holo-RbpA complex (**Figure 3.3**).

I screened the complex with T5 fork DNA, and one condition (0.25M LiSO4, 0.1 M Bis-Tris pH 6.0, 18% PEG 3350, 4 °C) yielded reproducible, chunky crystals ranging from 50-200 uM in size (**Figure 3.4b**). Despite the large size of the crystals and the extensive testing of various cryo conditions, the highest diffracting crystal diffracted only to 5.6Å (**Figure 3.4c**)—likely a result of the large solvent content of the crystals (>80%). Data on multiple crystals were collected at APS NE-CAT and combined for completeness, resulting in a 5.9Å data set.

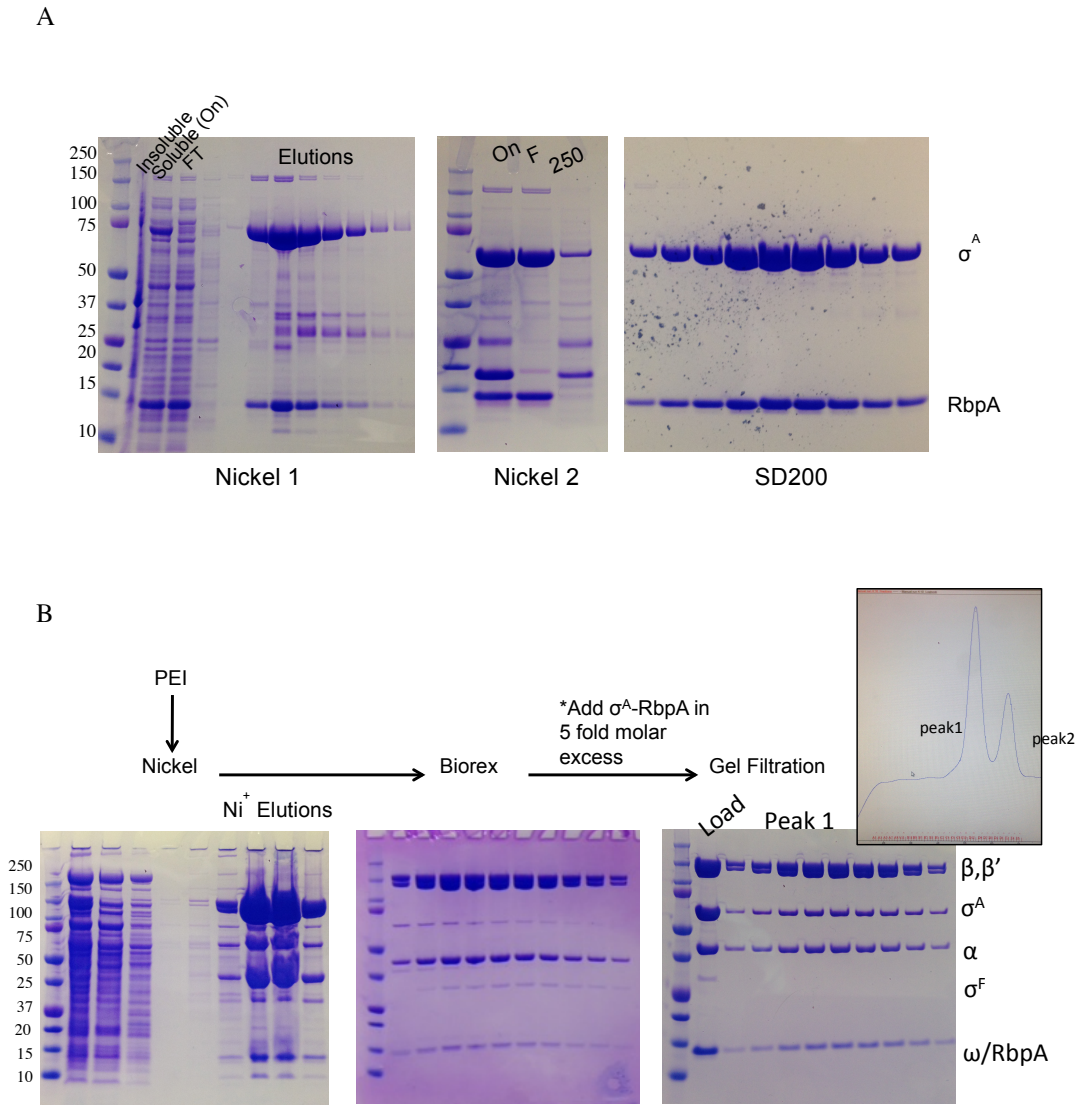


Figure 3.3: Purification of recombinant *Msm* σ^A/RbpA, endogenous *Msm* core RNAP and formation of *Msm* holo-RbpA. (A) his-SUMO-σ^A and RbpA from *Msm* are co-expressed in *Eco* from two separate plasmids in *Eco*. The proteins form a complex that is purified by nickel column, a subtractive nickel column (post cleavage of his-SUMO from σ^A by ULP1) and gel fitration (SD200). (B) Purification of endogenous RNAP from *rpoC*::*rpoC*-his10 *Msm* strain. Core RNAP was isolated by polymin P precipitation (PEI) and ammonium sulfate precipitation, followed by nickel and biorex columns. Contaminating σ^F was displaced by adding 5-fold excess of σ/RbpA (shown in (A)) prior to gel filtration by SD200, resulting in a highly pure, stoichiometric holo-RbpA complex.

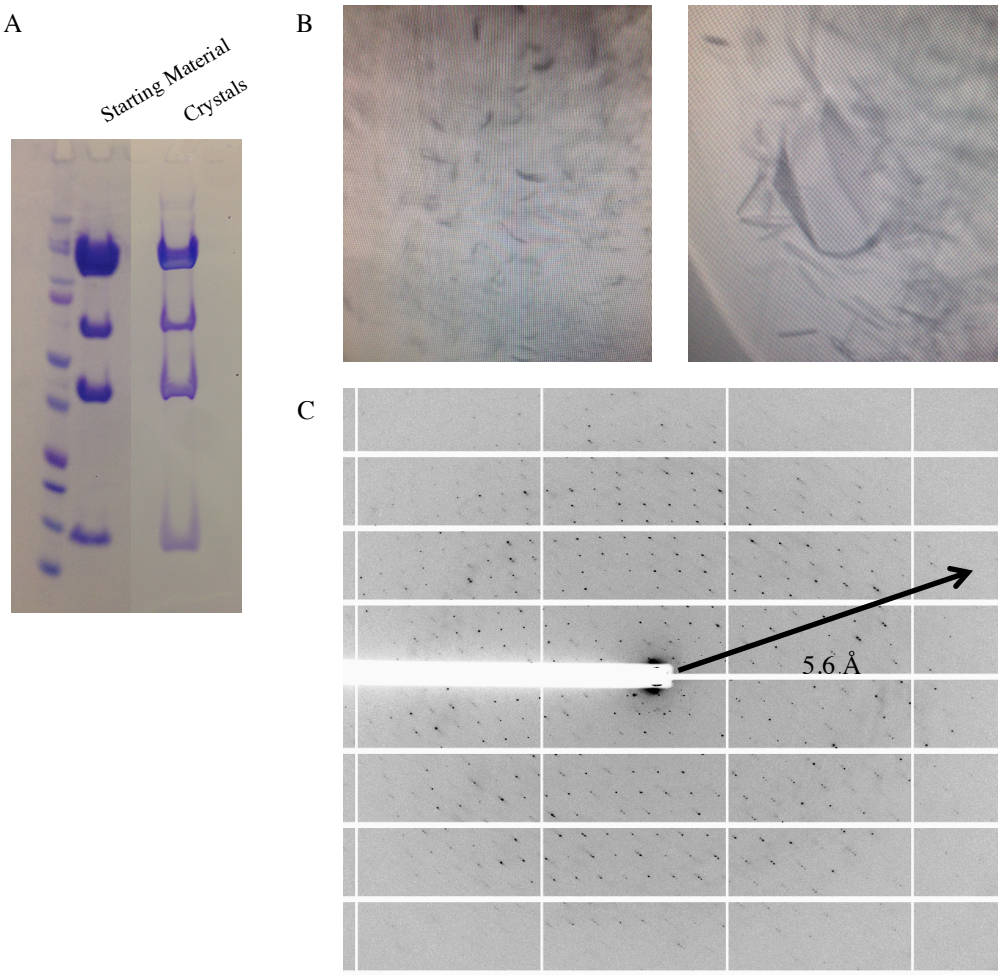


Figure 3.4: Crystallization of *Msm* holo-RbpA-T5. (A) Gel of crystals indicates they contain intact proteins. (B) Left: initial crystal hit. Right: optimized crystal and small rods. (C) Diffraction of crystals at APS NE-CAT. The best crystals diffracted to ~5.6 Å

The collected images were indexed, and the crystals had C2 spacegroup with very large unit cell dimensions (a=247.0 Å, b=212.93 Å, c=558.57 Å; α=γ=90, β=93.68). The crystals contained 4 molecules per unit cell, and crystal packing was driven by upstream DNA packing end-to-end with upstream DNA from a symmetry-related molecule (**Figure 3.5**). I was easily able to phase the data using molecular replacement with conserved regions of the *Taq* holoenzyme (4XLP (Bae et al., 2015), TFZ=21), and the map revealed the previously predicted lineage-specific α-helical insert in β' (β'i1, Figure 1A) found in Actinobacteria, Chloroflexi, Nitrospirae, and Mollicutes (Lane and Darst, 2010), indicating that the phasing of the maps was correct (**Figure 3.6a**).

While the electron density map confirms the location of the RbpA-SID on the holoenzyme predicted by our model, density for the rest of RbpA (CD or NTD) was absent, providing us with little new information (**Figure 3.6b**). Gels of the crystals indicated that RbpA had not undergone proteolysis within the drop (**Figure 3.4a**), so we hoped that a better diffracting crystal form would reveal the location of the rest of RbpA.

To drive packing of a different crystal form, we re-screened the *Msm* holo-RbpA complex with the T6 fork (1 bp shorter than the T5 fork) and T10 forks (1 bp longer than the T5 fork) (**Figure 3.2**). Promising microcrystals with the T10 fork appeared in similar conditions as the T5 fork crystals; however, they were different in morphology (**Figure 3.7a**). Optimization of crystallization conditions with the T10 fork (0.2 M $LiSO_4$, Bis-Tris pH 6.5, 16% PEG 3350) resulted in long (300 μm) rod-shaped crystals that diffracted better than 3Å (**Figure 3.7b**). Data from multiple crystals were collected and combined to generate a data set to 2.76Å.

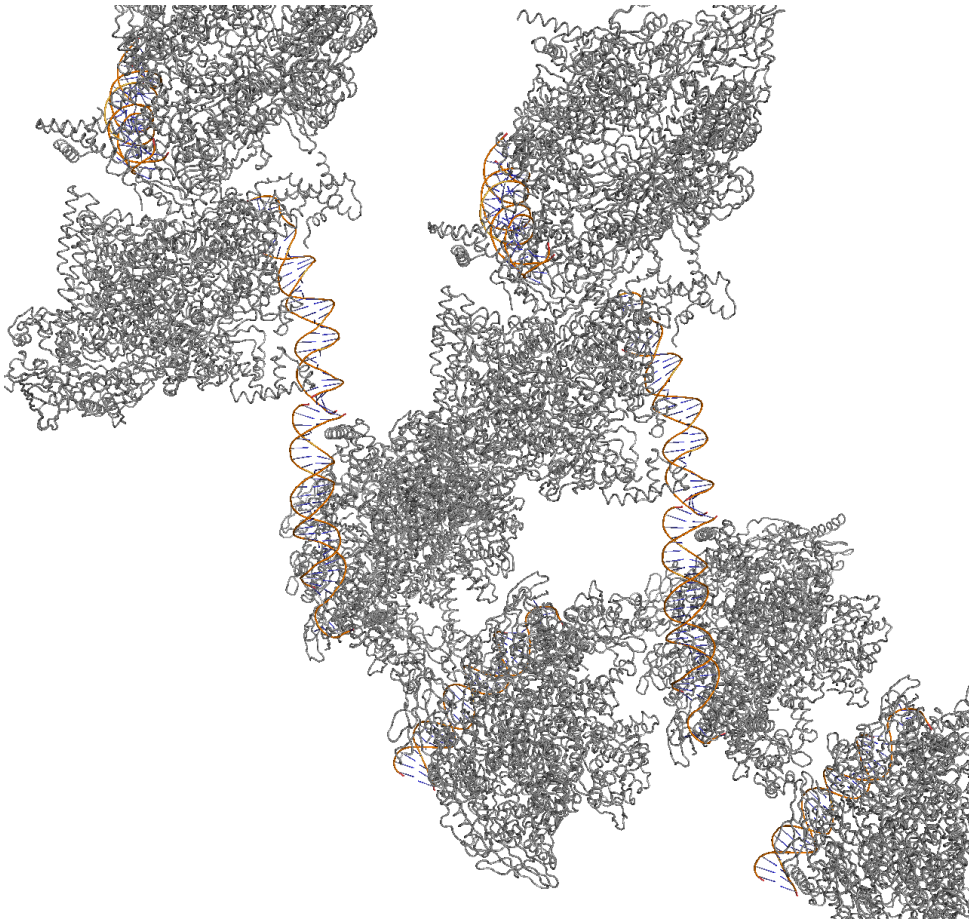
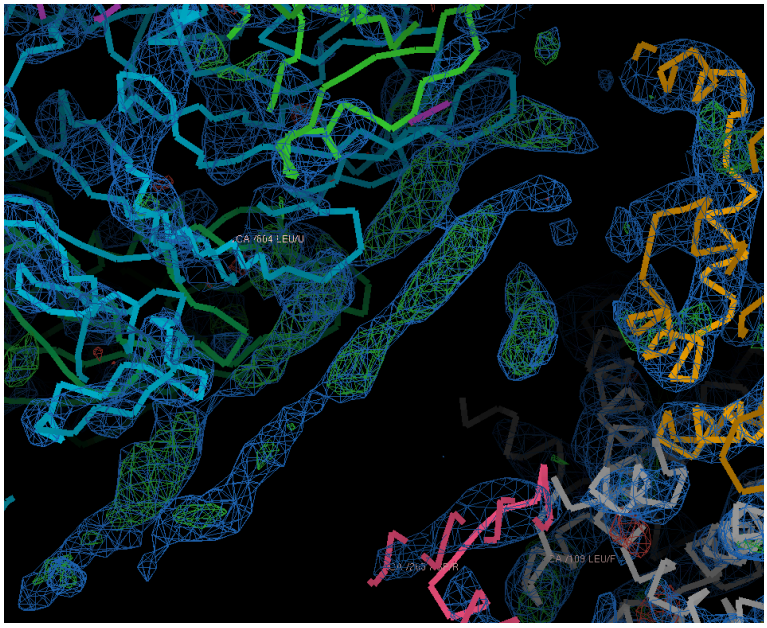


Figure 3.5 Crystal packing of *Msm* holo-RbpA-T5 was driven by upstream DNA packing end-to-end with upstream DNA from a symmetry molecule. RNAP colored in gray; DNA colored by element.

A



B

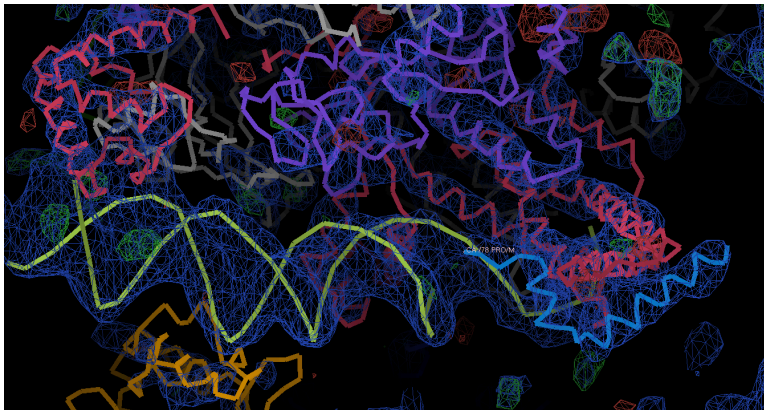


Figure 3.6: Electron density maps of *Msm* holo-RbpA bound to T5 fork at 5.9Å resolution. Electron density maps after initial rigid body refinement contoured to sigma of 1.0. 2FoFc map is in blue and contoured to sigma of 1.0, and Fo-Fc is in red/green and contoured to sigma of 3. (A) Electron density map reveals positive density for previously predicted lineage-specific a-helical insert in β' (β'i1) found in Actinobacteria, Chloroflexi, Nitrospirae, and Mollicutes. (B) Electron density map confirms the location of the RbpA-SID (colored in blue, bound to σ in pink) on an initiation complex; however, no density was present for the rest of RbpA.

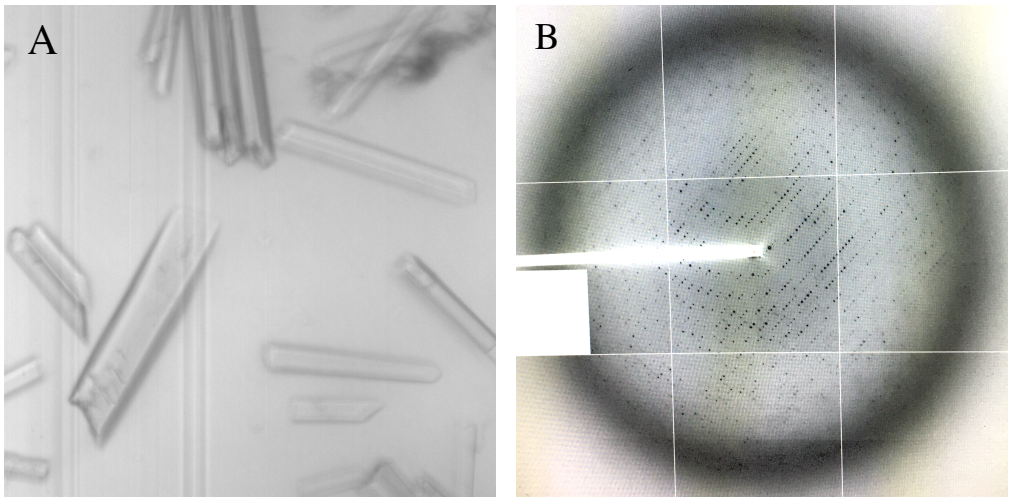


Figure 3.7: Crystallization of *Msm* holo-RbpA-T10 fork. (A) Optimized crystals (B) Diffraction at APS NE-CAT. Crystals diffracted to 2.7Å.

Use of the T10 fork DNA (**Figure 3.2**) disrupted end-to-end DNA packing seen with the T5 fork DNA (**Figure 3.8**), resulting in crystals in space group P2 with unit cell dimensions of 133.02Å x 161.63Å x 139.21Å (α=γ= 90°, β=107.99°) allowing for 1 complex per asymmetric unit and a solvent content of 60%. The decrease in solvent content was likely a determining factor for the increased resolution of this crystal form. Data were phased by a holoenzyme model containing the conserved region of *Taq* RNAP, fitted with the *Mtb* RbpA-SID-σ^{A}_{2} structure (Bae et al., 2015; Hubin et al., 2015; Chapter 2).

The density for the *Msm* holo-RbpA-T10-fork crystals was fitted with the *Msm* RNAP sequence by first replacing the *Taq* sequence with that of *Msm* and then performing rigid body refinements of mobile structural domains based on previous crystal structures solved in the Darst lab. After initial all-atom refinement, positive electron density for the RbpA-CD was clear, and we were able to fit the previously solved NMR structure of the *Mtb* RbpA-CD (residues 26-69) (Bortoluzzi et al., 2013) into the positive electron density, which aligned closely to the *Msm* RbpA-CD with an RMSD of 0.775 Å over 40 Cα atoms (**Figure 3.9**). The reiterative cycles of building (including building region 1.1 of σ, the RbpA CD, the LSI β'i1) and refinement were performed, and a final R/Free R of 0.2435/0.2816 was obtained (**Table 3.1**). No density was present for the unstructured, flexible RbpA-NTD, indicating that this region was disordered in our crystals.

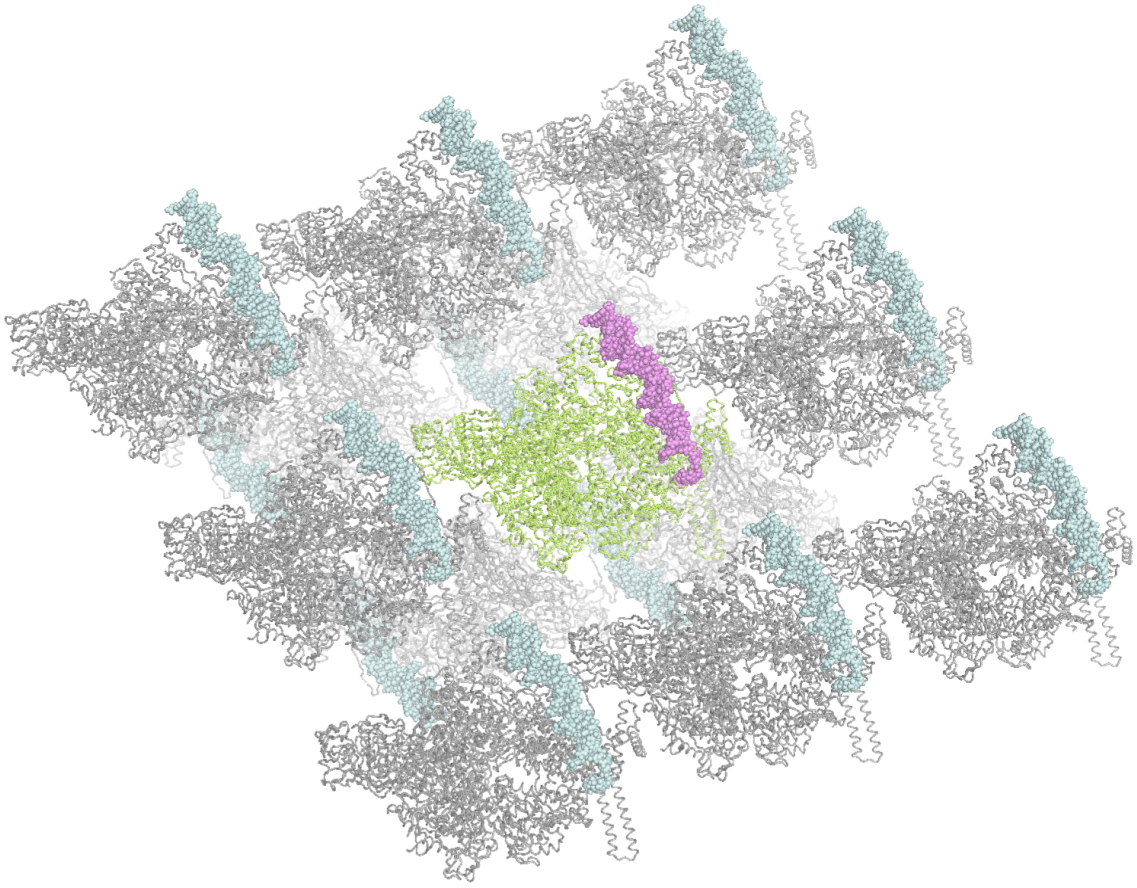


Figure 3.8: Crystal packing of *Msm* holo-RbpA-T10 fork crystals. RNAP symmetry molecules shown in gray with DNA in blue.

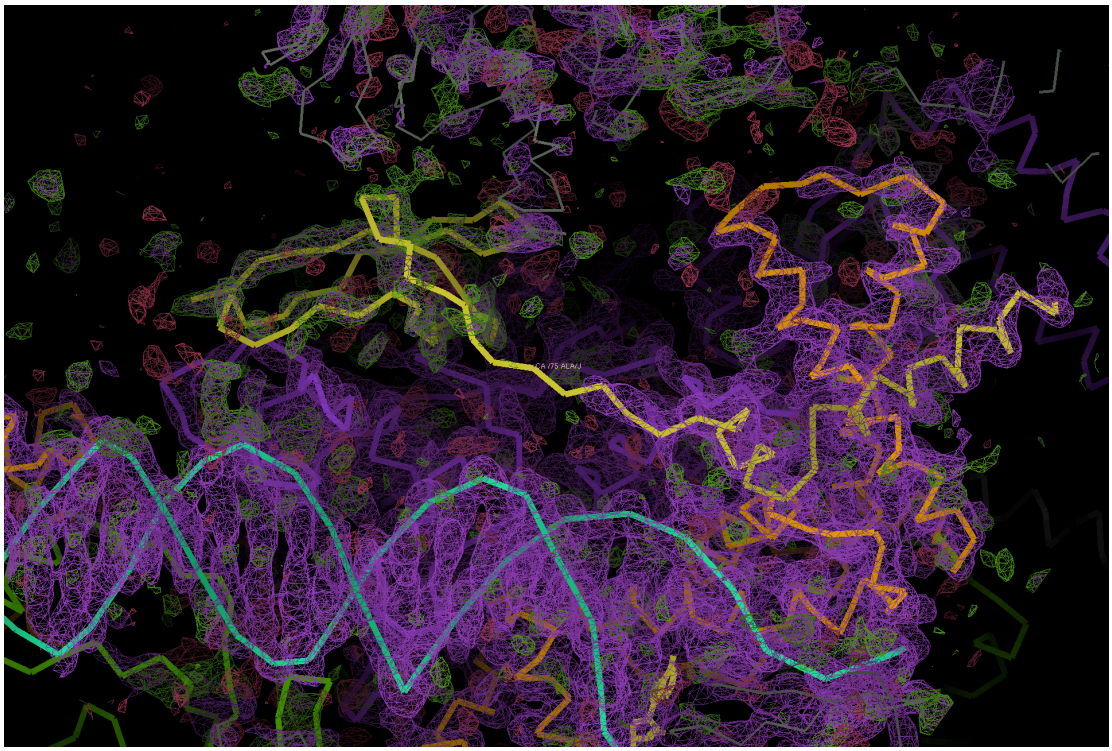


Figure 3.9: Electron density map of *Msm* holo-RbpA-T10 after initial refinement. 2FoFc map is purple and contoured to 1.0 sigma. FoFc map is red/green and contoured to 2.5 sigma. RbpA (yellow) is bound to σ (orange) near DNA (teal). Map shows clear positive green density (FoFc) corresponding to the RbpA-CD.

Table 3.1. Crystallographic statistics.

	Msm holo/us-fork/RbpA
Data collection	
Space group	$P2_1$
Combined datasets	4
Cell dimensions	
a (Å)	133.012
b (Å)	161.633
c (Å)	139.211
Wavelength (Å)	0.97918
Resolution (Å)	51.99 – 2.76 (2.859 – 2.76)[a]
Total reflections	2,329,541 (175,429)
Unique reflections	143,776 (13,955)
Multiplicity	16.2 (12.3)
Completeness (%)	99 (100)
$\langle I \rangle / \sigma I$	22.39 (0.77)
Wilson B-factor (Å^2)	78.20
R_{merge}[b]	0.2343 (4.816)
R_{meas}[b]	0.2417 (5.021)
R_{pim}[b]	0.059 (1.564)
CC1/2[c]	0.998 (0.214)
CC*[c]	1 (0.594)
Refinement	
R_{work} / R_{free}	0.2435/0.2816 (0.4446/0.4513)
CC_{work}/CC_{free}[c]	0.950/0.932 (0.446/0.418)
No. atoms	26,600
Macromolecule	26,401
Ligand/ion	87
Water	112
Protein residues	3,332
B-factors	
Macromolecules	87.02
Ligand/ions/water	105.83
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.64
Clashscore	36.12
Ramachandran favored (%)	95
Ramachandran outliers (%)	0.27

[a] Values in parentheses are for highest-resolution shell.

3.3 The X-ray crystal structure of *Msm* holo-RbpA-DNA reveals that RbpA CD binds to the β' ZBD and zipper

The structure of the *Msm* initiation complex bound to RbpA (**Figure 3.10, Table 3.1**) reveals that the RbpA–CD binds to the clamp region of β', which is consistent with our previous structural modeling. The RbpA–CD, comprising four β-strands, contacts the β' zipper and zinc-binding domain (ZBD) of the RNAP (**Figure 3.11**), two highly conserved structural elements that make up part of the clamp domain (Yuzenkova et al., 2011). The interactions between the CD and the zipper and ZBD are both polar and non-polar and bury a surface area of 615 $Å^2$. An alignment of RbpA from 856 bacterial genomes indicates that although the CD itself is only ~55% conserved, the residues that contact the ZBD and zipper are ~80% conserved (**Table 3.2**). A representative alignment of RbpA across bacterial species is shown in (**Figure 3.12**).

Consistent with structures of *Taq* and *Eco* initiation complexes, the zipper region in our *Msm* RNAP structure contacts DNA within the spacer region. Specifically, residue R37 forms ionic interactions with the phosphate group at the -18 position of the nt-strand (distance=~3.3Å). A study has shown that this interaction between the zipper and DNA may play a role in recognition of promoters with specific "z-elements" (bases -18 to -12) (Yuzenkova et al., 2011). Of note, the CD does not interact with the region of the zipper that contacts the DNA, nor does the CD itself contact DNA.

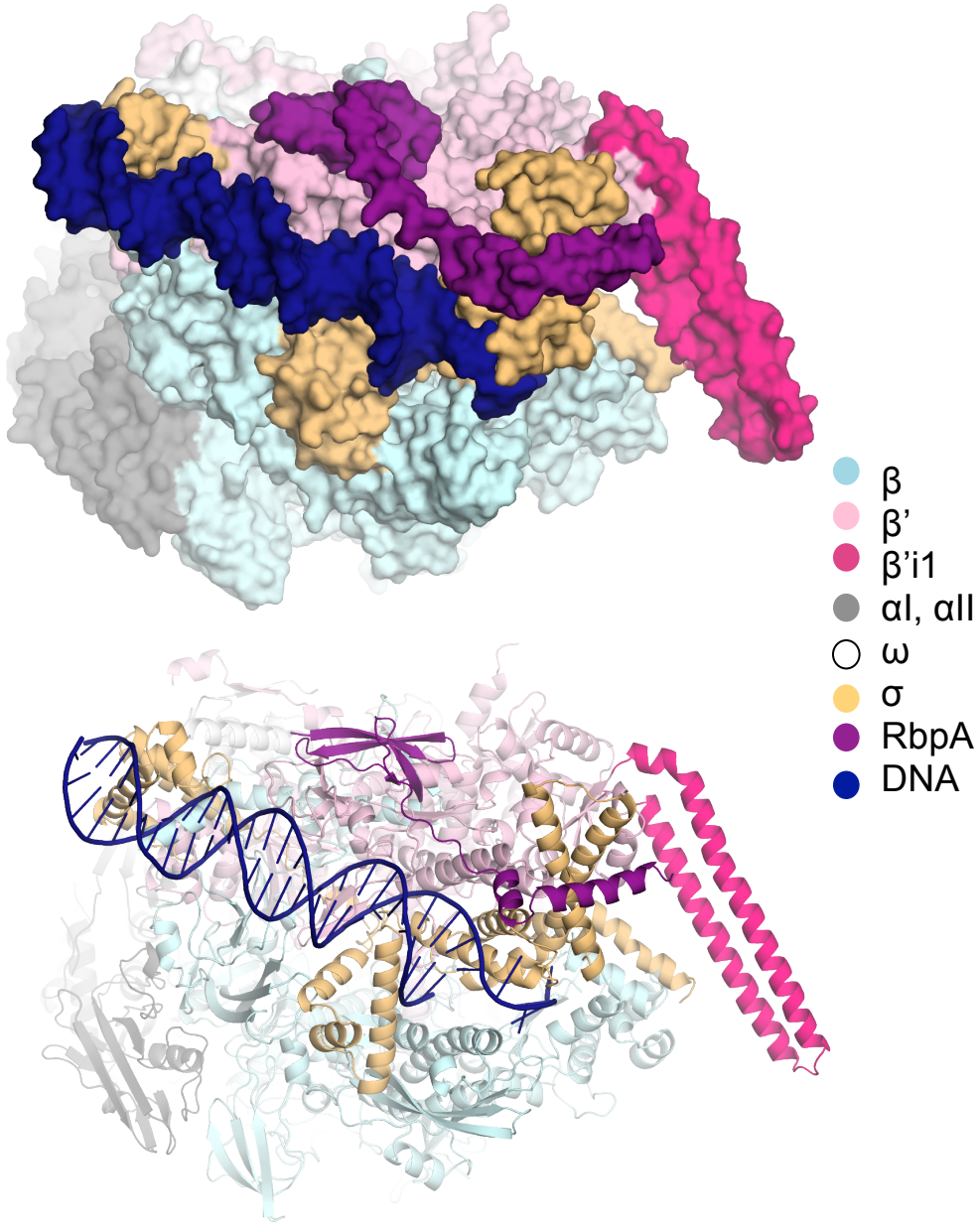


Figure 3.10: Crystal structure of the *Msm* σ^A-containing initiation complex bound to RbpA. Proteins and DNA are colored as indicated. Top: Surface. Bottom: Secondary structure, depicted in ribbon.

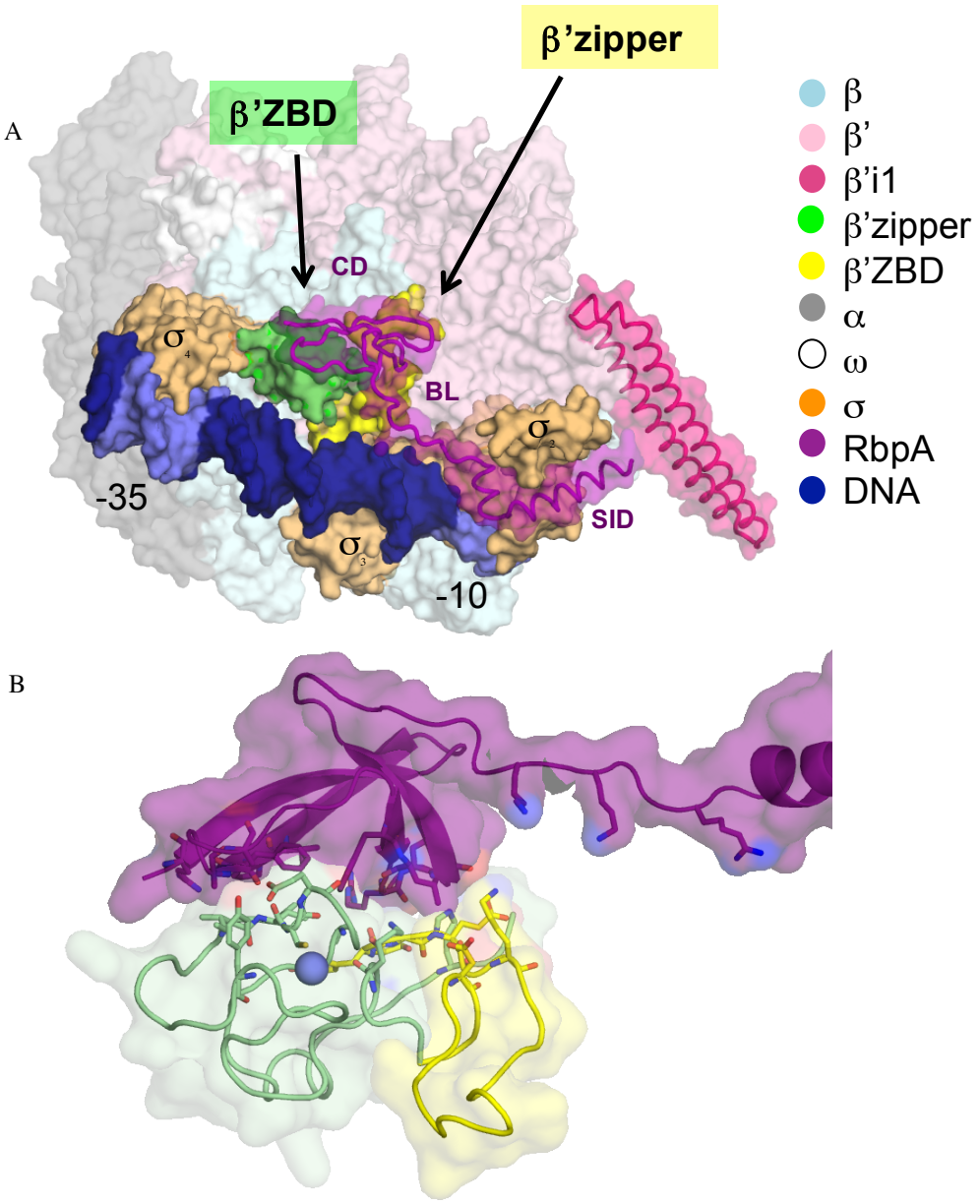


Figure 3.11: RbpA CD interacts with the β' zipper and β' ZBD. Top: RbpA (purple) interacts with both σ (tan) and with two regions on β': the zipper (yellow) and the ZBD (green). R79, K76 and K74 in the BL, and residues involved in interaction between the CD are depicted in stick. DNA is colored by element with the -10 and -35 regions in black**.** (B) RbpA's interaction with the zipper and ZBD (colored as in above) are highlighted in stick.

Table 3.2: Summary of interactions between *Msm* RbpA–CD and RNAP-β' subunit

RbpA	β'	SC/MC	Interaction
R27	I73	SC-SC	H
	G72	SC-MC	P
V42	I74	SC-SC	H
P43	I74	MC-MC, SC-SC	P, H
F44	I74	SC-MC	H
	E76	SC-SC, SC-MC	H
	I73	MC-SC	H
	C75	SC-MC	H
A45	E76	MC-SC	P
	I73	MC-SC, SC-SC	H, H
	Y65	SC-SC	H
A48	E76	SC-SC	H
	Y65	SC-SC	H
W54	G79	SC-MC	H
	I74	SC-SC, SC-MC	H
	E76	SC-MC	H
L55	K50	MC-SC, SC-SC	P, H
	D44	SC-SC	H
	K29	SC-SC	H
R57	Y25	MC-MC, SC-MC	H, H
	G26	MC-MC, MC-MC, SC-MC	H, P, H
	E27	MC-MC	P
	H94	MC-SC	P
	S24	SC-MC	H, P
	R21	SC-MC	P
	N22	SC-MC	P
N58	H94	MC-SC, MC-SC	H, P
G59	E27	MC-SC	H
	K29	MC-SC, MC-SC	H, P

H=hydrophobic, P=polar. No ionic interactions observed. MC- main chain interaction, SC-side chain interaction. RbpA residue conservation is indicated by color: Red=absolutely conserved; Orange=Functionally conserved or more than 80% conserved; Green= more than 60% conserved; blue=more than 50% conserved; Purple= less than 50% conserved

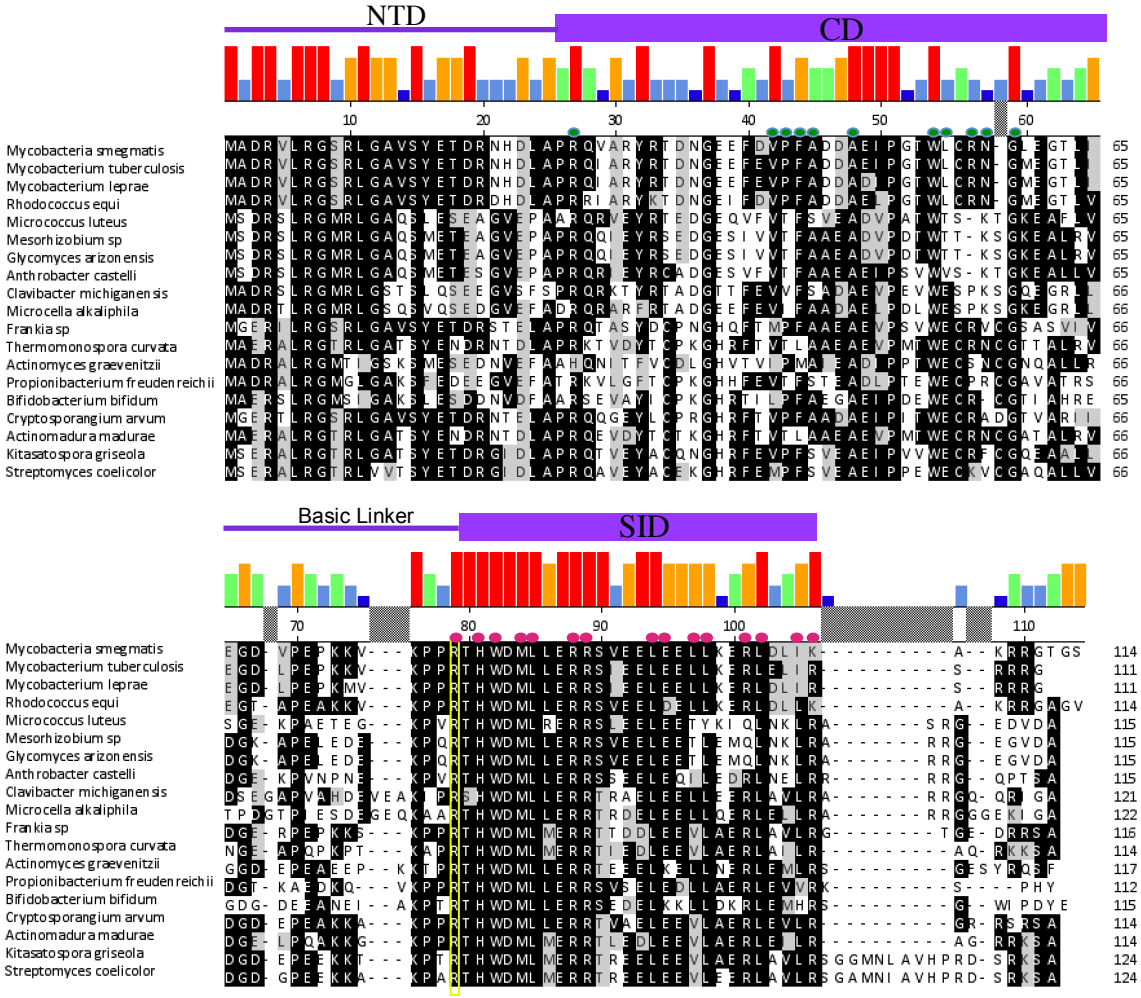


Figure 3.12: Representative alignment of RbpA from 856 bacterial species. RbpA residues that contact the β' are marked by blue dots, and residues that contact the σ^{A}_{2} are marked by pink dots. Absolutely conserved R79 is boxed in yellow. Numbering corresponds to *Msm* RbpA.

The ZBD comprises a zinc finger in which 4 cysteine residues coordinate with a Zn^{2+} ion. It is one of two Zn binding elements in prokaryotic RNAPs, neither of which is conserved in eukaryotes. In the *Msm* holoenzyme structure, the ZBD is in close proximity to DNA at the -20 bp but does not directly interact with DNA. It does, however, interact directly with the zipper and with σ_4, which both contact DNA.

The RbpA-SID's association with σ closely aligns with the *Mtb* σ^A RbpA-SID (RMSD=0.451 over 152 Cα atoms), and the extensive contacts between the SID and σ_2 (conserved and NCR) match those described in Chapter 2.

3.4 Structural evidence that conserved arginine 79 (R79) residue interacts with DNA upstream of the transcription bubble

Based on our previous structural model, absolutely conserved arginine residue (RbpA-R79) of the BL was predicted to bind to the phosphate backbone at the upstream edge of the transcription bubble. The residue has been found to be essential for both RbpA's activation *in vitro* and modulation of holo-DNA binding in anisotropy assays and required as well for normal growth of *Streptomyces coelicolor in vivo* (Hubin et al., 2015). The structure of the *Msm* RNAP initiation complex with RbpA confirms that RbpA-R79 forms extensive ionic interactions with the phosphate group between the -13 and -14 bases of the non-template strand (**Figure 3.13**). Conserved RbpA lysine residues, K76 and K74, do not directly contact DNA in our structure; they come within 5.5 Å of the phosphate backbone. The lysine residues possibly stabilize interactions with the DNA, contributing a favorable electrostatic interface to the phosphate backbone of the DNA or directly interacting with DNA during a transient initiation complex.

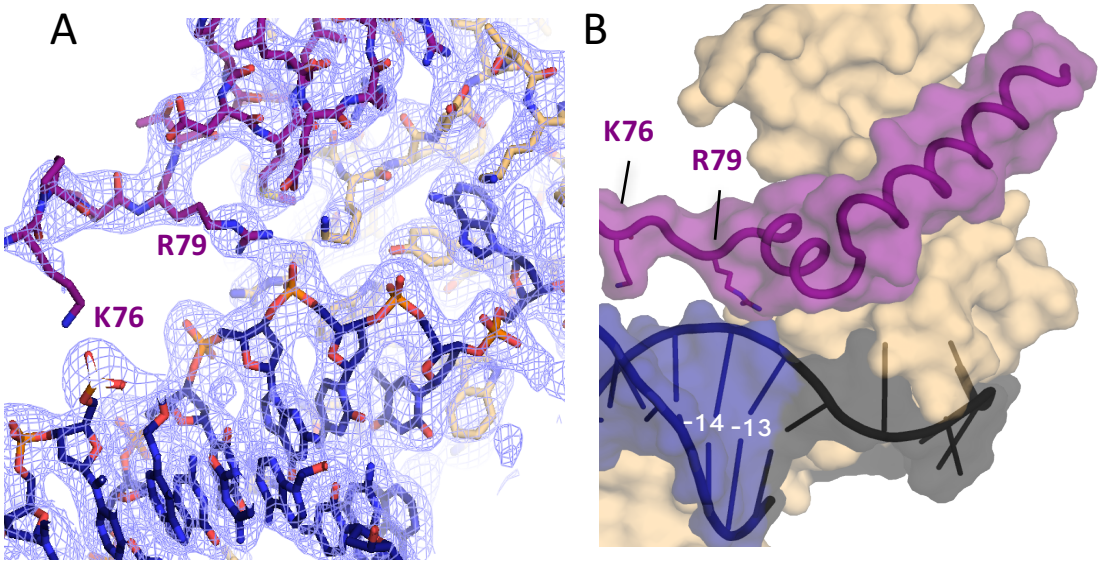


Figure 3.13: Conserved residue RbpA-R79 interacts with duplex DNA upstream of the transcription bubble. (A) Electron density map contoured to 1σ. RbpA (colored in purple) and DNA are shown in sticks. The map shows R79 directly contacts DNA between the -13 and -14 bases. Density indicates that K76 comes within 5.5 A of the DNA backbone. (B) Depiction of interactions described in (A). DNA is colored in blue with the -10 element colored in black.

An alignment of RbpA from over 1200 species (data not shown) reveals that some bacteria, including *Sco* species, carry a second copy of RbpA (RbpA2), which shares only ~30% sequence identity with RbpA. Functional studies of RbpA2 have not been reported; however, the alignment revealed that in RbpA2 proteins, the residue corresponding to R79 is a lysine in ~80% of the RbpA2 proteins and an arginine in the other ~20%. In RbpA, R79 is 100% conserved. Taken together, this indicates that in both RbpA and RbpA2, this residue is absolutely conserved to have a basic side chain group that functions to contact the acidic phosphate backbone of the promoter DNA upstream of the -10 element.

3.5 Discussion

Whether or not RbpA interacts with core RNAP was previously unknown. Two studies predicted RbpA to bind two disparate regions on the β subunit; however, these binding determinants were completely incompatible with our structure-based model discussed in the previous chapter. The *Msm* holoenzyme bound to the T-10 fork DNA and RbpA reveals that the CD binds to the clamp domain of the β' subunit, specifically interacting with the ZBD and the zipper region. The function of the CD, which is not necessary for transcription activation, remains unknown. I explore the significance of the CD's interactions with the zipper and ZBD in the context of biochemical data in the discussion section of Chapter 4.

Despite the high resolution of the *Msm* holo-T10-RbpA structure, no density was present for the RbpA-NTD, likely because this region is highly flexible. Given the disordered nature of the NTD, this 26-residue region could occupy a wide area. It is still

unknown whether or not this region interacts with RNAP or, alternatively, binds another regulatory factor. Future studies are necessary to provide both a functional and structural understanding of this region in the context of an initiation complex. Functional studies investigating domain function of RbpA, including the RbpA-NTD, will be described in Chapters 4 and 5.

During refinement of the *Msm* holo-T10-RbpA structure, unexpected density appeared corresponding to an α-CTD from a symmetry-related molecule interacting with an A/T-rich region of the fork DNA from the -25 to the -30 bp (**Figure 3.2**). Using an *Eco* α-CTD structure to fit into the density, we built the *Msm* α-CTD (**Figure 3.14a).** Although canonical A/T rich UP elements are not thought to be present in mycobacterial promoters (Arnvig et al., 2005), this structure shows that the α-CTD of the mycobacterial RNAP is still able to associate with an A/T rich regions. It is possible that this interaction is simply due to the conserved nature of the α-CTD between bacteria; however, it should not be ruled out that certain mycobacterial promoters are regulated by the α-CTD interaction with upstream DNA and possible UP-elements. To date, only a few studies have investigated the significance of upstream DNA in the mycobacterial transcription system.

Another notable element of the *Msm* holo-T10-RbpA structure is the presence of a helix corresponding to part of the N-terminal region 1.1 of σ ($\sigma_{1.1}$), a region that occupies the space that the downstream DNA needs to be placed for open complex formation (Bae et al., 2013). The precise location where $\sigma_{1.1}$ is displaced upon RPo formation was previously unknown, as this region is not present in any published structures of RNAP initiation complexes. Previous biochemical studies predicted this region to be located

near the β protrusion (also referred to as β-lobe 2) (Meckler et al., 2002). We see density for a 15-amino acid helical segment of $\sigma_{1.1}$ (residues 145-160 in mycobacteria, which have insert N-terminal to $\sigma_{1.1}$) in a similar general area predicted by Meckler et al. Notably, $\sigma_{1.1}$ is in very close contact to LSI β'i1(**Figure 3.14b**). The significance of the proximity of $\sigma_{1.1}$ to LSIs will be discussed in Chapter 6.

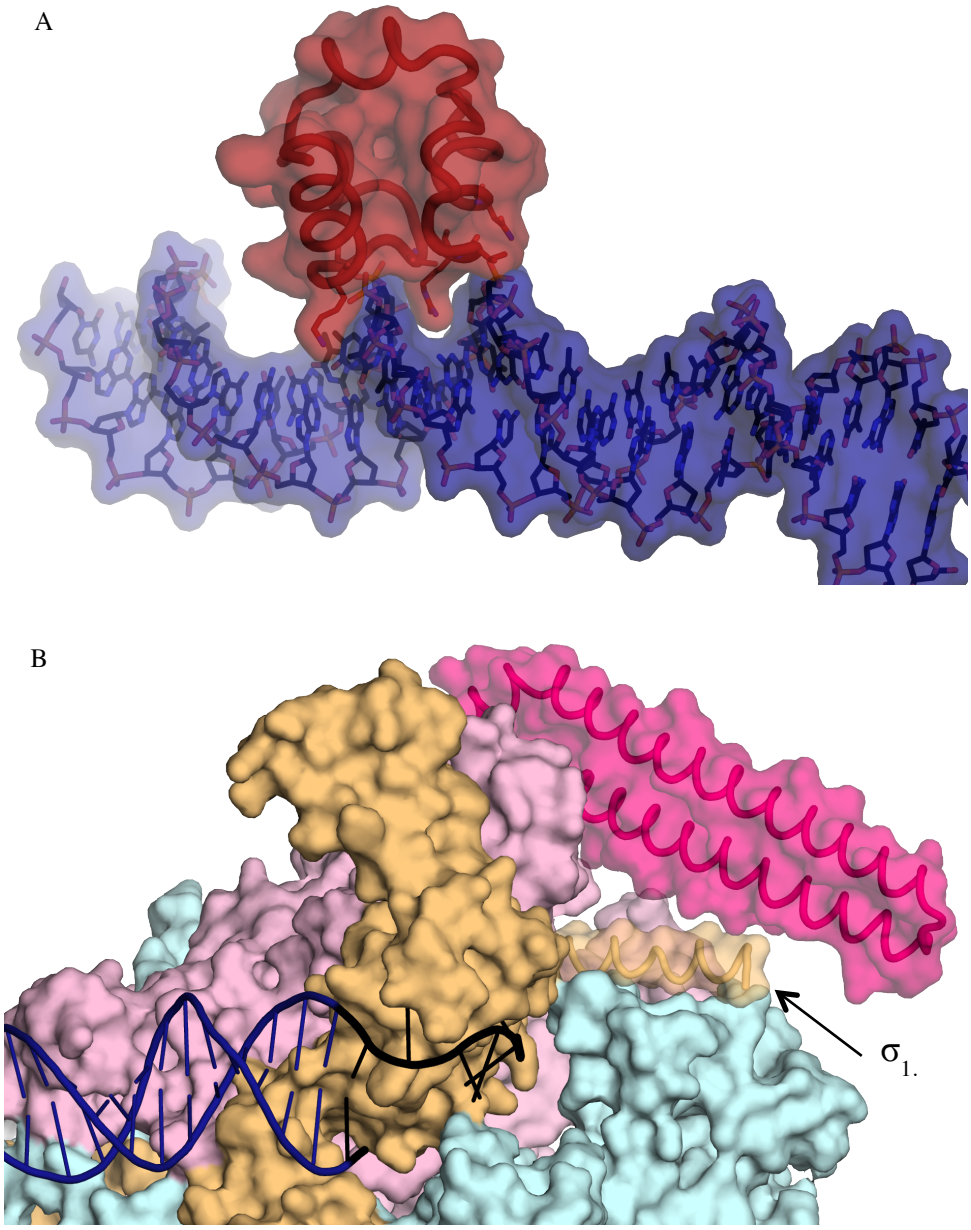


Figure 3.14: Additional elements of *Msm* holo-RbpA-T10 structure. (A) *Msm* α-CTD bound to A/T rich spacer region (-25 to -30) of T10 fork DNA. α-CTD is shown in red; DNA in blue. (B) Part of region 1.1 of σ is revealed in the *Msm* RNAP structure to be located near the protrusion (β-lobe 2) and coming in close contact to LSI β'i1. DNA is shown in blue, σ in orange, β in light blue, β' in LSI β'i1 in fuschia.

Chapter 3 Acknowledgements:

Allison Fay of Mike Glickman's lab produced an *Msm* strain containing a his_{10}-tag fused to the native chromosomal *rpoC* gene, which was used for endogenous purification of *Msm* RNAP. Freezing of crystals was performed with the assistance of Dr. Brian Bae of the Darst Lab. Building and refinement of the *Msm* holo-RbpA-T10 structure was completed with the assistance of Dr. Elizabeth Campbell of the Darst Lab. Many of the crystallization trials described in this chapter were performed with the assistance of Catherine Xu, a summer undergraduate student in the Darst Lab.

Chapter 4:

Functional Analysis of RbpA and its Relationship to CarD

The crystal structures presented in the previous chapters provide the structural architecture of RbpA bound to an initiation complex: the SID tightly interacts with the NCR and conserved parts of domain 2 of the housekeeping σ-factor; the BL contacts promoter DNA upstream of the transcription bubble, pointing to BL-DNA contacts as a basis for transcription activation; and the CD interacts with the ZBD and zipper on the RNAP β' subunit. However, the roles of the CD and disordered NTD (which was not revealed in our *Msm* holo-DNA-RbpA structure) remain unknown. In this chapter, I further explore the contribution of RbpA's individual domains.

Recent work has shown that RbpA and CarD can function cooperatively to stabilize RPo formation (Rammohan et al., 2016). Here I confirm that these two factors have a synergistic effect on transcription initiation and also investigate RbpA domain function in the context of CarD. To assess RbpA domain function *in vivo* in *Msm,* we collaborated with Michael Glickman's laboratory at Sloan Kettering.

4.1 RbpA and CarD have a synergistic effect in abortive initiation assays

Modeling of *Tth* CarD from the recently published initiation complex (Bae et al., 2015a) on our *Msm* holo-RbpA-DNA structure confirmed that the initiation complex can accommodate both RbpA and CarD simultaneously (**Figure. 4.1**). Thus, I tested RbpA

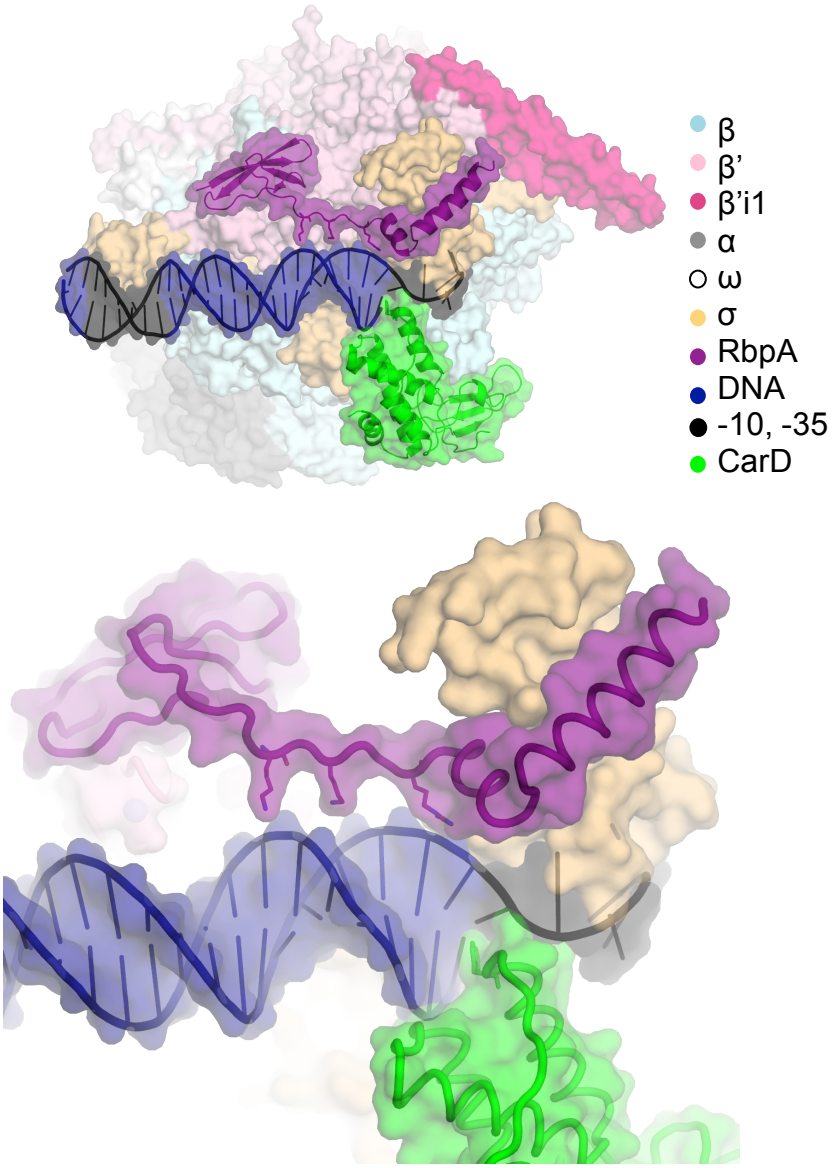


Figure 4.1: *Msm* holoenzyme can accommodate both RbpA and CarD. Top: *Msm* holo-RbpA-DNA structure modeled with *Tth* CarD (Bae et al., 2015). Proteins and DNA are colored as indicated. Bottom: zoomed in view of (A) to highlight Rbpa and CarD situated across from one another at the upstream edge of the DNA.

and CarD's relationship *in vitro* transcription assays on various promoters (see **Figure 4.4; Appendix 1** for promoter sequences).

When RbpA and CarD were added together, I repeatedly saw activation greater than when either factor was added alone. Throughout this section I will refer to this effect as: a *combined effect* if it is less than additive; an *additive effect* if it is additive; and a *synergistic effect* if the effect is more than additive. While a combined or additive effect simply confirms that the two proteins do not compete with each other, a synergistic effect indicates that the presence of one factor enhances the activity of the other. Such synergy was obvious in abortive initiation assays on the VapB promoter when competitor DNA was present (**Figure 4.2**). These results are consistent with a recent report by Rammohan et al., 2016, showing that RbpA and CarD act cooperatively on RPo formation, and synergy between the two factors indicates they may function together *in vivo*.

4.2 The RbpA domain function in *in vitro* transcription assays

Knowing that $RbpA^{SID\text{-}BL}$ is sufficient for full activation, I tested whether full length RbpA is necessary for a synergy with CarD. I evaluated activation of WT *Mbo* RNAP by four RbpA constructs with and without CarD: FL RbpA (**Figure 4.3a**); RbpA lacking an NTD ($RbpA^{\Delta NTD}$) (**Figure 4.3b**); RbpA composed of just the SID and BL and lacking both the RCD and NTD ($RbpA^{SID\text{-}BL}$) (**Figure 4.3c**); and RbpA with conserved arginine residue, R79, mutated to alanine ($RbpA^{R79A}$). The assays were performed as multi round abortive initiation assays (no competitor present) on AP3 and VapB promoters, as well as a modified AP3 promoter with a non-consensus -35 ($AP3^{\Delta\text{-}35}$) (**Figure 4.3a-c;** sequences in **Appendix 1**).

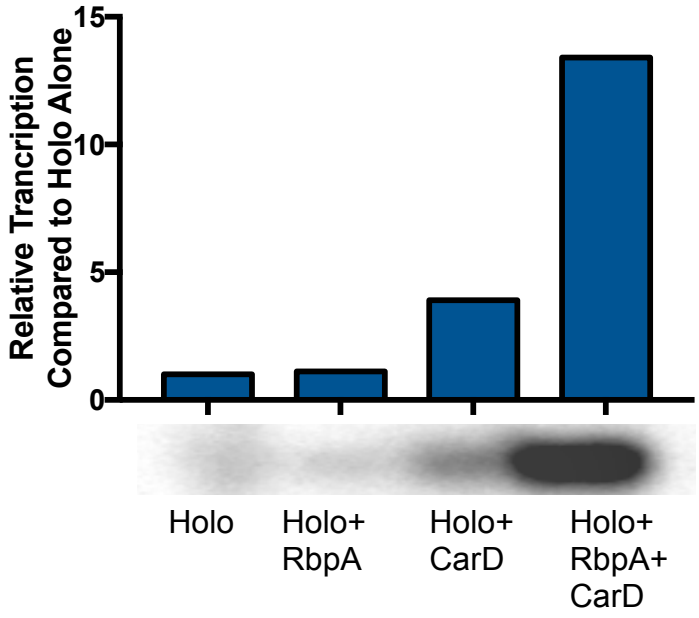


Figure 4.2: RbpA and CarD have a synergistic effect on transcription. Abortive initiation assay was performed on VapB with competitor DNA present. Activity was measured by production of [32P]-labeled abortive transcript (ApUpG). Graphical representation is normalized to abortive transcript produced by holoenzyme without additional transcription factors. Reactions were performed using *Mbo* holo RNAP (50 nM) with and without transcription factors RbpA and CarD (2 μM). Reactions were initiated with an ApU dinucleotide primer and [32P]-labeled GTP. When added together, the two factors had a synergistic effect on transcription.

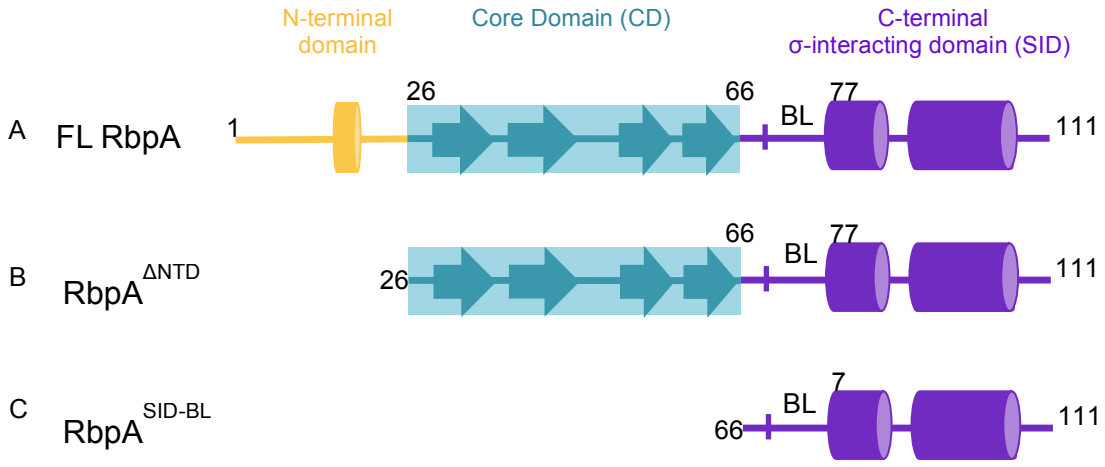


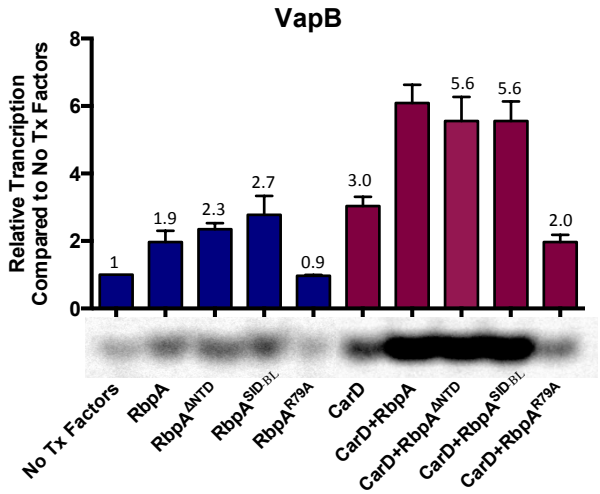
Figure 4.3: Schematic of RbpA protein truncations used to study RbpA domain function. (A) Full length RbpA (*Mtb* residues 1-111), (B) RbpA$^{\Delta NTD}$ (*Mtb* residues 26-111), (C) RbpA^{SID-BL} (*Mtb* residues 66-111) comprises just the sigma interacting domain (SID) and basic linker (BL).

On VapB, in the absence of CarD, full-length RbpA, RbpA$^{\Delta NTD}$, and RbpA$^{SID\text{-}BL}$ all activated WT *Mbo* RNAP to a similar extent, with the RbpA$^{SID\text{-}BL}$ activating slightly (but not significantly) better than the other constructs (**Figure 4.4a**). As expected, the RbpA^{R79A} showed no activation. Notably, the N-terminal RbpA truncations had a similar synergistic effect with CarD as full-length RbpA had. There was also a curious and reproducible effect of RbpA^{R79A} repressing transcription, especially noticeable when CarD was present (**Figure 4.4a**).

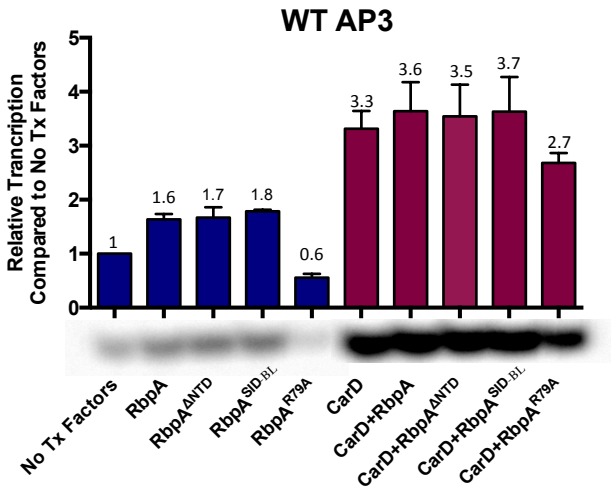
On AP3, RbpA and the RbpA derivatives activated transcription similarly as they had on VapB, with the decrease in transcription by RbpA^{R79A} also quite noticeable even in the absence of CarD (**Figure 4.4b**). We observed that the effect between RbpA and CarD was more evident on VapB than on AP3, and because VapB lacks a -35 element like the majority of mycobacterial promoters (Cortes et al., 2013), we hypothesized that synergy between the two factors might be more obvious on promoters lacking a -35 element. I therefore performed the same abortive assays on an AP3-based promoter with an anti-consensus -35 element, based on *Eco* consensus promoters (Shultzaberger et al., 2007). On the mutant AP3 promoter, transcription was decreased more than 10-fold compared to WT, and transcription without CarD was barely measurable (**Figure 4.4c**). However, the effect of FL RbpA and the RbpA truncations could be measured by a synergistic effect with CarD. Interestingly, in the presence of CarD, RbpA$^{SID\text{-}BL}$ activated significantly more than WT RbpA, whereas RbpA$^{\Delta NTD}$ activated significantly less than WT. The R79A inhibitory effect, both with and without CarD, again was noticeable. Further examination of synergy between CarD and the RbpA derivatives is discussed in section 4.3.

Figure 4.4 (Following page): The SID-BL is sufficient for full activation while RbpA R79A represses initiation. Transcription activation by various RbpA truncations and RbpA R79A mutant with and without the presence of CarD. Activity was measured by production of [32P]-labeled abortive transcript (ApUpG for vapB and GpUpU for AP3) in multi-round abortive initiation assays. Graphical representation is normalized to abortive transcript produced by holoenzyme without additional transcription factors. No competitor was added. (A) VapB promoter; $RbpA^{FL}$, RbpA-$^{\Delta NTD}$ and RbpA-$^{SID\text{-}BL}$ have a combined effect on transcription activation with CarD. Reactions were initiated with ApU and [32P]-labeled GTP. Error bars represent standard error of 3 experiments. (B)WT AP3 promoter. Reactions were initiated with GpU and [32P]-labeled UTP. Error bars represent standard error of 2 experiments. (C) AP3 promoter lacking a -35 element; $RbpA^{SID\text{-}BL}$ has a greater combined effect with CarD than $RbpA^{WT}$ and $RbpA^{\Delta NTD}$. Reactions were initiated with GpU and [32P]-labeled UTP. Error bars represent standard error of 3 experiments

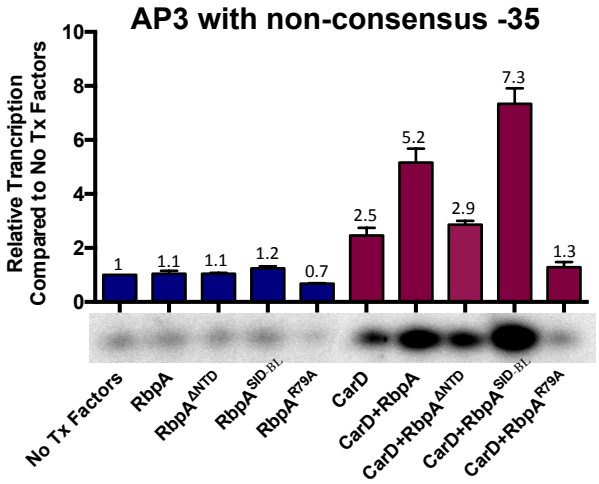
A


VapB
Relative Trancription Compared to No Tx Factors
0
2
4
6
8
1
1.9
2.3
2.7
0.9
3.0
5.6
5.6
2.0
No Tx Factors
RbpA
RbpA ΔNTD
RbpA SID BL
RbpA R79A
CarD
CarD+RbpA
CarD+RbpA ΔNTD
CarD+RbpA SID BL
CarD+RbpA R79A

B


WT AP3
Relative Trancription Compared to No Tx Factors
0
1
2
3
4
5
1
1.6
1.7
1.8
0.6
3.3
3.6
3.5
3.7
2.7
No Tx Factors
RbpA
RbpA ΔNTD
RbpA SID BL
RbpA R79A
CarD
CarD+RbpA
CarD+RbpA ΔNTD
CarD+RbpA SID BL
CarD+RbpA R79A

C


AP3 with non-consensus -35
Relative Trancription Compared to No Tx Factors
0
2
4
6
8
10
1
1.1
1.1
1.2
0.7
2.5
5.2
2.9
7.3
1.3
No Tx Factors
RbpA
RbpA ΔNTD
RbpA SID BL
RbpA R79A
CarD
CarD+RbpA
CarD+RbpA ΔNTD
CarD+RbpA SID BL
CarD+RbpA R79A

I tested whether full-length RbpA is necessary for full activation using *in vitro* run off assays, in which, unlike abortive initiation assays, the RNAP must escape the promoter to transcribe a full-length RNA product. In run-off assays on both AP3 and VapB, the $RbpA^{SID-BL}$ was the most active construct on both promoters (**Figure 4.5**).

The transcription assays described in this section show that not only is the RbpA-SID-BL sufficient for full activation on the three promoters tested, but in some cases, it is *more* active than full-length RbpA (**Figure 4.5**) suggesting that the CD may have a role in *inhibiting* transcription. Further evidence of the CD inhibiting transcription is the effect of the R79A mutant, which represses transcription on WT AP3 and mutant AP3 promoters and represses transcription in the context of CarD on all three promoters (**Figure 4.4**). Because the R79A mutant essentially removes the activating function of the BL, yet still allows the CD to bind core RNAP, inhibition by the R79 mutant, together with the observed increased activity of $RbpA^{SID-BL}$ in run-off assays, indicates that the CD represses transcription initiation.

4.3 RbpA and CarD work synergistically to extend the half-life of RPo

Studies indicate that the main function of CarD is to stabilize RPo, which is made particularly evident by its effect of dramatically increasing the lifetime of competitor-resistant initiation complexes (Bae et al., 2015a; Davis et al., 2015). Because mycobacterial RNAP forms very unstable open complexes compared to *Eco* (Davis et al., 2014), I examined whether RbpA also plays a role in stabilizing RPo using abortive initiation half-life assays (**Figure 4.6**).

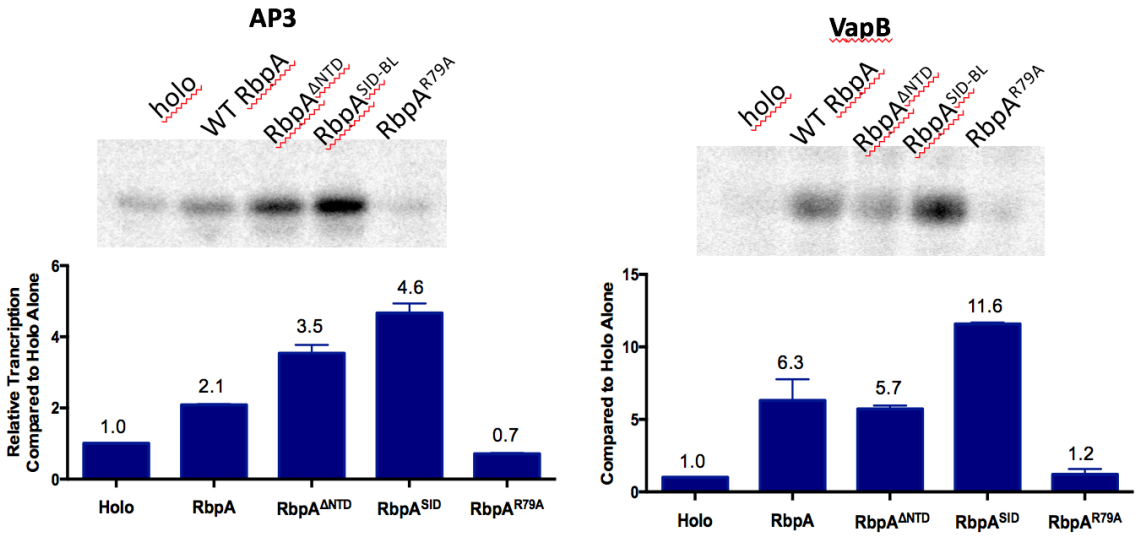


Figure 4.5: The SID activates transcription more than FL RbpA in run off transcription assays on both AP3 and VapB promoters. Activity measured by production of [32P]-labeled run-off transcript. Graphical representation is relative to transcript produced by holoenzyme without RbpA. Error bars represent standard error of two experiments.

Interestingly, RbpA alone had no impact on the half-life of RPo compared to holo; however, when added with CarD, it magnifies CarD's effect on half-life almost 3-fold (**Figure 4.6**). I evaluated whether the RbpA-NTD or CD is necessary for this synergistic effect on half-life and found that synergy between $RbpA^{\Delta NTD}$ and CarD is similar to that of full-length RbpA and CarD. I also found that $RbpA^{SID\text{-}BL}$ magnifies CarD's effect on the half-life to an even greater extent, which clearly shows that only the SID-BL is necessary for synergy between the two proteins. The $RbpA^{R79A}$ destabilized the half-life significantly, which, taken together with the greater effect of the $RbpA^{SID\text{-}BL}$, corroborates my data that the CD represses transcription initiation (**Figure 4.4**). The half-life of RPo is not only a function of bubble stability, but is also affected by the stability of initiation intermediates (Tsodikov et al., 1999). In Chapter 5, I will describe the kinetic basis for RbpA and CarD's combined effect on the lifetime of initiation complexes.

4.4 The NTD and CD are necessary for normal growth *in vivo*

The transcription assays presented in this section show that the NTD and CD are dispensable for transcription stimulation *in vivo*. However, our collaborators in Mike Glickman's group found that these domains are necessary for normal growth of *Msm in vivo*. Glickman's group discovered that *Msm* ΔRbpA cells were not viable, confirming that RbpA is essential in *Msm*. Integrative expression of WT RbpA in the cells gave a normal doubling time of 2.6 hours (**Table 4.1**). Although the integration of $RbpA^{\Delta NTD}$ and $RbpA^{SID\text{-}BL}$ produced viable cells, the double time increased to 3.1 and 4.1 hours, respectively. Whereas $RbpA^{\Delta NTD}$ exhibited similar growth phenotype to WT, the cells harboring the $RbpA^{SID\text{-}BL}$ had severe growth and chromosomal segregation abnormalities,

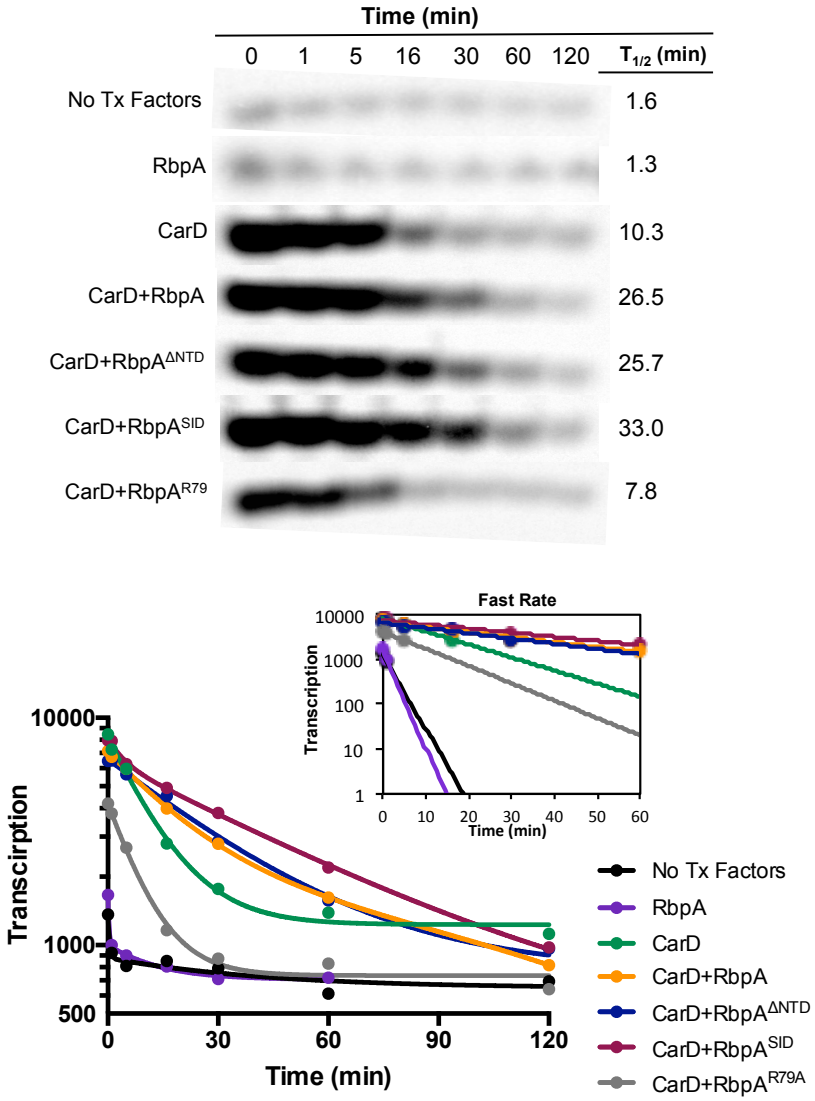


Figure 4.6: RbpA enhances CarD's effect on RPo formation. (A) Lifetime of AP3 promoter complexes were measured by production of [32P]-labeled abortive transcript at times after the addition of a large excess of competitor promoter DNA trap. Calculated half-lives shown to the right. (B) Transcript production in (A) quantified and plotted as a double exponential and used to calculate half-live values. Top plot removes secondary decay.

indicating the CD is critical for normal growth and cell division (**Figure 4.8**). The above results indicate that the NTD and CD have an important function in the cell.

4.5 Discussion:

The transcription assay data that I presented in this chapter not only show that the RbpA SID-BL is fully sufficient for transcription activation but also suggest that the RbpA CD *represses* transcription initiation. However, the finding that the CD and NTD are required for normal growth in *Msm in vivo* (**Figure 4.7**, **Table 4.1**) indicates that these regions do have important functions that are currently unknown.

It has been suggested that RbpA may function to promote holoenzyme assembly of σ^A and σ^B-containing holoenzymes by acting as a bridge between σ and core (Flenti et al., 2016; Hu et al., 2012). The biochemical data presented in this chapter clearly indicate that this is not RbpA's primary function for transcription activation *in vitro*—evident from increased activity of the $RbpA^{SID\text{-}BL}$. It is possible, however, that full-length RbpA plays a role in promoting the assembly of σ^A and σ^B-containing holoenzymes in the context of the crowded cellular environment, where many σ factors (such as σ^F) compete for limiting core RNAP. This function, however, does not explain the curious effect of the CD repressing transcription initiation *in vitro*.

The CD contacting the zipper and the ZBD of the β' subunit (as described in detail in Chapter 3) may provide clues to the function of this domain. The zipper region has been shown to stabilize promoter complexes and has a role in recognition of promoters with a specific "z-element" between the -22 and -18 positions (Yuzenkova et al., 2011). However, structural data (Bae et al., 2015) reveal only interactions between

Table 4.1: *Msm* cells harboring RbpA truncations are viable but slow-growing

ΔrbpA	**Doubling time (hours)**
attB::kan	Not viable
attB::rbpA (wt) kan	2.61
attB::rbpA(28-114aa) kan	3.13
attB::rbpA(72-114aa) kan	4.10

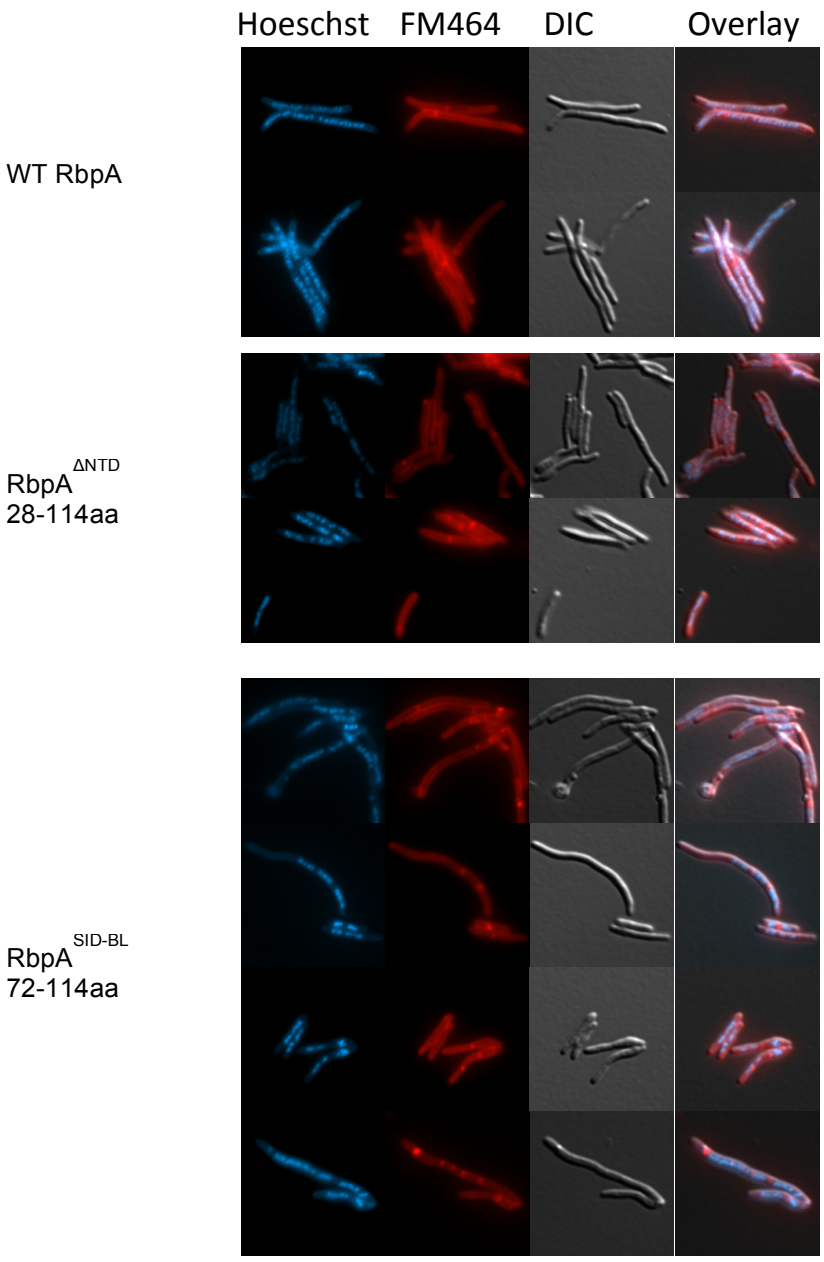


Figure 4.8: Growth phenotypes of *Msm* cells harboring RbpA truncation mutants. Cells were stained for nucleic acid (Hoechst) and membranes (FM464). RbpA$^{\Delta NTD}$ exhibited similar growth phenotype to that of WT (top and middle panels); however, RbpA^{SID-BL} (bottom panels) had severe growth and chromosomal segregation abnormalities, as indicated by the presence of budding cells without nuclei.

the zipper and the DNA phosphate backbone, so it is unclear whether these interactions involve specific contacts to DNA bases, or whether, instead, the interactions are simply specific to spacer regions with particular conformation. It was proposed that promoters lacking the -35 may rely on zipper interactions with the z-element or that this interaction could be an element of gene control (Yuzenkova et al., 2011). It's possible that the RbpA-CD allosterically modulates the zipper's ability to bind to promoters with specific z-elements that are not present in AP3 or VapB. Again, however, this function does not explain the CD effect of repressing transcription initiation.

Researchers have also found that the zipper plays a role in regulation of σ-dependent pausing (described in the introduction), as *Eco* RNAP containing mutated or deleted zippers had *less* pausing but had a *longer* pause half-life (Yuzenkova et al., 2011). Little is known about the function of the ZBD, but it has been shown to be required for promoter melting, indicating an involvement in transcription initiation (Young et al., 2004; Yuzenkova et al., 2011). The ZBD has also been shown, along with the zipper, to play a role in allosterically regulating termination (Epshtein et al., 2007).

The roles of the zipper and ZBD in termination and the zipper's regulation of σ-dependent pausing indicate that both of these elements modulate the stability of RNAP promoter complexes. It is possible that the function of the RbpA-CD is to bind these elements and to facilitate destabilization of RNAP contacts with the promoter DNA, product RNA, or both, in a mechanism analogous to termination or anti-pausing. Such a function could explain the CD's inhibition of transcription *in vitro*, as its presence could also destabilize the RNAP-promoter contacts required for transcription. We currently don't fully understand the CD's negative effect on transcription but acknowledge that a

role in modulating the zipper's effect on σ-dependent pausing would allow for fine tuning for gene expression, depending on the promoter.

A recent study reported that RbpA and CarD have a cooperative effect on RPo formation and that the presence of one factor increased the activity of the other (Rammohan et al., 2016). However, this study could not say definitively whether or not this effect was because of a direct interaction between the two proteins, or an allosteric mechanism mediated by changes in the RNAP-promoter complex during RPo formation. Although the *Msm* holo-RbpA-DNA structure rules out a direct interaction between RbpA and CarD, the structure cannot preclude an interaction between the NTD and CarD. The transcription assays I describe in this chapter unambiguously show that neither the NTD nor CD are necessary for a combined effect between the two factors, ruling out that a direct interaction between the NTD and CarD is the cause for the cooperative effect between RbpA and CarD. The exception to this finding was the abortive initiation assay on AP3 with a non-consensus -35, where, curiously, the combined effect between $RbpA^{\Delta NTD}$ and CarD was repeatedly not observed (**Figure 4.4c**). Further experiments are required to understand this finding, but it is possible that the NTD's role may be promoter-dependent.

The synergy between CarD and RbpA in *in vitro* transcription assays indicates that these two factors could work together to stabilize promoter complexes in the cell. We propose that CarD and RbpA are part of the general transcription machinery in mycobacteria. Our *in vitro* and *in vivo* results suggest that RbpA and CarD are both general transcription factors (GTFs) and make up part of the general transcription machinery during normal, exponential growth. Their non-canonical mechanisms function

to compensate for the unstable RNAP-promoter complexes formed in mycobacteria compared to those formed in *Eco.*

Chapter Acknowledgements:

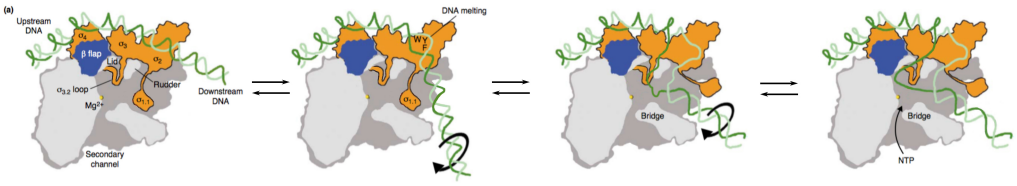
The *in vivo* work I describe in this chapter was performed in collaboration with Allison Fay of the Glickman lab.

Chapter 5:

Kinetic Analysis of RbpA and CarD Activation

Open complex (RPo) formation is a multi-step pathway, and discrete, 'kinetically significant' intermediates in this pathway have been previously identified using *Eco* RNAP holoenzyme on specific *Eco* promoters (Saecker et al., 2011) (**Figure 5.1**). Early studies described RNAP-promoter intermediates on *lac*UV5 and identified a closed complex (RPc) intermediate where the RNAP is bound to duplex DNA prior to any melting. Additional studies on this promoter (Buc and McClure, 1985) revealed that, in addition to an RPc intermediate, a second intermediate, RPi could be identified; thus RPo formation on *lac*UV5 could be characterized by at least a three-step mechanism. A similar three-step mechanism has been identified on *Eco* promoters T7A1 (Sclavi et al., 2005) and *rrnB* P1 (Rao et al., 1994; Rutherford et al., 2009) (**Figure 5.1**). The second intermediate is hypothesized to contain partially melted DNA and a 90° bend at the -10 element that places the downstream duplex DNA across the entrance of, but not in, the β,β' channel (Ruff et al., 2015).

Studies of the *Eco* promoter λP_R show that RPc does not build up to a significant extent in this case; the first kinetically significant intermediate, called *I1*, is thought to be similar to *lac*UV RPi (**Figure 5.1**). Kinetic characterizations of λP_R (Saecker et al., 2011) have also identified an additional intermediate, *I2*, which contains a nearly full transcription bubble with DNA in the active site channel, but the single-stranded DNA within the transcription bubble is thought to not be positioned properly as in the final RPo.



*lac*UV5[a]	PRc	RPi		RPo
λP_R[b]		*I1*	*I2*	RPo
T7 A1[c]	A	B		C
rrnB P1[d]	RPc	RPi		RPo

Figure 5.1: Structural models and names of intermediates in the mechanism of RPo formation on various promoters. Images adapted from Murakami and Darst, 2003.
[a] Studitsky et al., 2001
[b] Saecker et al., 2011
[c] Sclavi et al., 2005
[d] Rutherford et al., 2009

$KMnO_4$ footprinting shows that the transition from *I2* to RPo involves repositioning of the discriminator region of the nt-strand into the active site cleft (Ruff et al., 2015).

To date, the majority of experiments studying these intermediates have been performed using non-real time assays (such as foot-printing assays, EMSA, filter-binding and abortive transcription assays), in which RPo formation must be stopped and then measured. In contrast, a recently reported fluorescence assay (Ko and Heyduk, 2014) allows for the detection of RPo formation in real time and does not depend on the use of competitors, such as heparin. This assay, which I refer to as the 'Heyduk' assay, measures RPo formation based on changes in the fluorescence of a Cy3 fluorophore attached to promoter DNA at the +2 position of the non-template strand (**Figure 5.2**). In the original study of Ko and Heyduk (2014), the authors showed that the Cy3 label does not perturb the process of RPo formation and used the assay to report on *Eco* RPo formation on λP_R. However, the assay was not used to analyze RNAP-promoter intermediates that occur during RPo formation. Fluorescence by Cy3 is very sensitive to environmental changes; a large increase in fluorescence occurs as the dye is transferred from a polar to a hydrophobic environment (Toutchkine et al., 2003). The proposed conformational changes that occur during RPo formation (**Figure 5.1**) would place the dye in different environments; thus it is likely that this assay would report on intermediate species in which RNAP has bound to the promoter, but RPo has not fully formed. Further, Ko and Heyduk (2014) found that fluorescence progress curves over time were minimally fit by a double exponential non-linear regressive model, rather than a single exponential non-linear regressive model, suggesting the presence of an

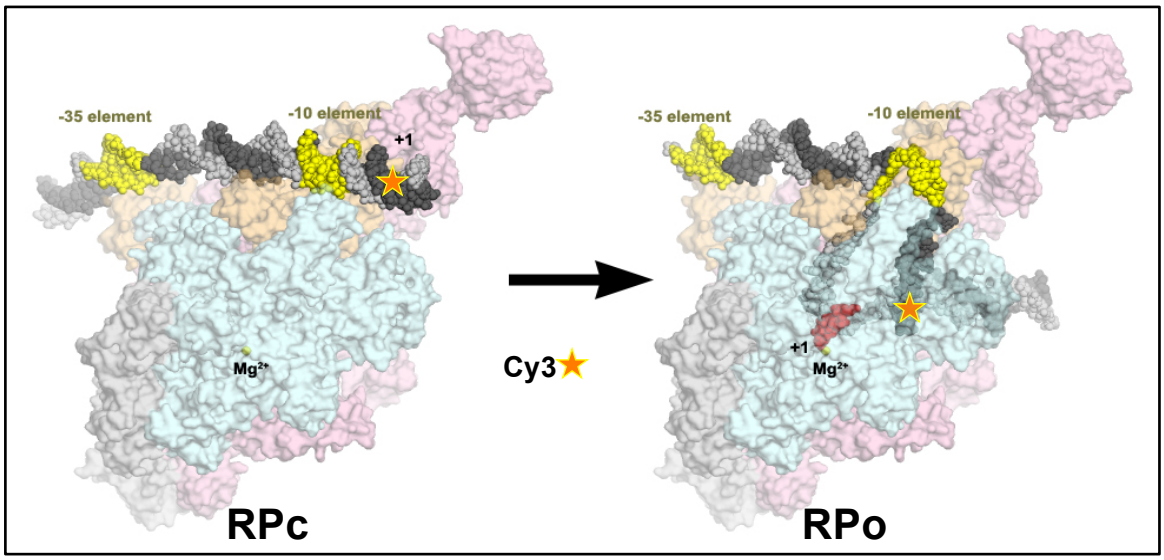


Figure 5.2: Fluorescence-based 'Heyduk' assay to measure RPo formation. (Ko & Heyduk. 2014)**.** Promoter DNA is labeled at the +2 position. The changes in environment upon RNAP binding and melting of DNA increases the fluorescent signal of the probe.

intermediate species population. Thus, this assay may be well suited for detecting intermediates on the RPo formation pathway in a relatively non-disturbing manner.

The half-life ($t_{1/2}$) of RPo is a function of both forward and reverse kinetic steps in this pathway (Tsodikov et al., 1999). Since RbpA does not extend the lifetime of RPo by itself but has a cooperative effect with CarD (Chapter 4; **Figure 4.6**), the two transcription factors act by different mechanisms. The Heyduk assay was recently used to evaluate CarD and RbpA's effect on RPo formation on Cy3 labeled *Mtb* AP3 (Rammohan et al., 2016; 2015), showing that CarD and RbpA indeed have a cooperative effect on RPo formation. In these studies, however, individual rate constants relating to formation of initiation complex intermediates were not reported, and therefore a mechanism for RbpA and CarD's cooperative effect could not be determined. Using the Heyduk assay and Kintek Explorer (Johnson et al., 2009) to fit our data, we were able to estimate individual rates for a multi-step mechanism of RPo formation (described in detail in Section 5.2) and determine how RbpA and CarD influence these rates. The data presented in this chapter provided us with a mechanistic model for the cooperative effect of RbpA and CarD in transcription initiation.

5.1 Monitoring RPo formation on the *Mtb* AP3 promoter using the Heyduk assay

I performed the Heyduk assay using a KinTek stopped-flow instrument that monitored the change in fluorescence of Cy3-labeled AP3 promoter DNA (-61 to +53 with the Cy3 label at the +2 position on the nt strand; **Figure 5.3**) upon mixing with either *Mbo* holo (with and without CarD and RbpA) or *Eco* holo over a range of RNAP concentrations (**Table 5.1**). Equal volumes of DNA and protein were rapidly mixed, and

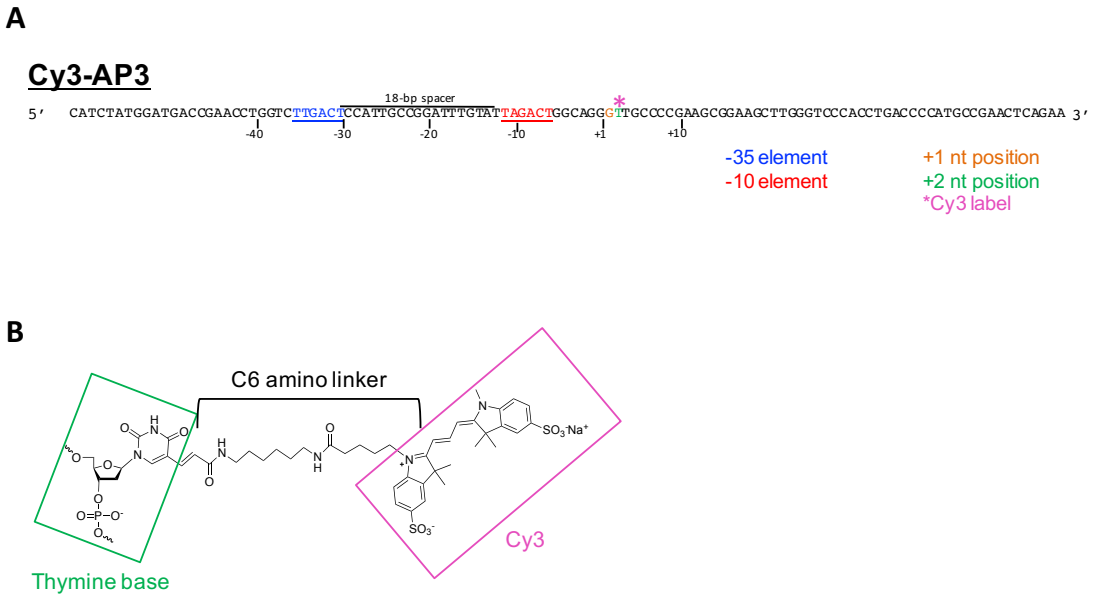


Figure 5.3: Cy3-AP3 DNA used for Heyduk Assay. (A) Non-template strand sequence of Cy3-AP3 DNA (-61 to +53). Elements are colored as indicated. Cy3 label (pink) is attached to the +2 thymine base (green). An 18 base-pair spacer region separates the -10 and -35 element. (B) Chemical modification: Cy3 (pink) is attached to the C6 position of the Thymine base (green) through a linker with two amino groups.

Table 5.1: Concentrations of protein and DNA for Heyduk assay experiment with Cy3-AP3

Sample	[Cy3-AP3]	[RNAP] range	[RbpA]	[CarD]
Eco holo	1 nM	5-400 nM	-	-
Mbo holo	1 nM	3-400 nM	-	-
Mbo holo + RbpA	1 nM	3-400 nM	5 μM	-
Mbo holo + CarD	1 nM	3-200 nM	-	10 μM
Mbo holo + RbpA+ CarD	1 nM	3-200 nM	5 μM	10 μM

fluorescence was monitored over time. A final concentration of 1 nM promoter DNA was used for each reaction, and the highest concentration of RNAP went to saturation. When RbpA and CarD were added, they were kept at constant saturating concentrations (5 μM and 10 μM, respectively) (Table 5.1) so that we could be sure that when the factors were present, the assay did not report on any RNAP species that were not bound by factors. This contrasts with the experiments reported by Rammohan, et al. (2015 and 2016), which involved keeping RNAP concentration constant while titrating in the factors. Although their method could report on cooperativity between RbpA and CarD, having both unbound and bound RNAP populations complicates the calculation of individual rate constants, as RPo formation would involve multiple species of RNAP (with and without factor).

I compared RPo formation by *Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+CarD, Mbo holo+RbpA+CarD and *Eco* holo. It was clear, even prior to quantitative analysis, that the behavior of *Mbo* RNAP was dramatically different from that of *Eco* RNAP. For instance, for each reaction, fluorescence was monitored until the curve leveled off (**Figure 5.4**). For *Mbo* holo, this took ~200s for an intermediate RNAP concentration (100 nM), while for *Eco* holo, this took ~30s (**Figure 5.4a-b**). Additionally, while even the lowest *Eco* RNAP concentrations eventually went to saturation (i.e. reaching the same maximum fluorescence as the high RNAP concentrations), the reactions with *Mbo* holo at low or intermediate concentrations of RNAP leveled off at very different fluorescence values. This indicates that there is profound reversibility of the *Mbo* RPo species, and that at low or intermediate concentrations of RNAP, *Mbo* RPo is in equilibrium with populated intermediates and/or free RNAP and promoter DNA. This is consistent with biochemical

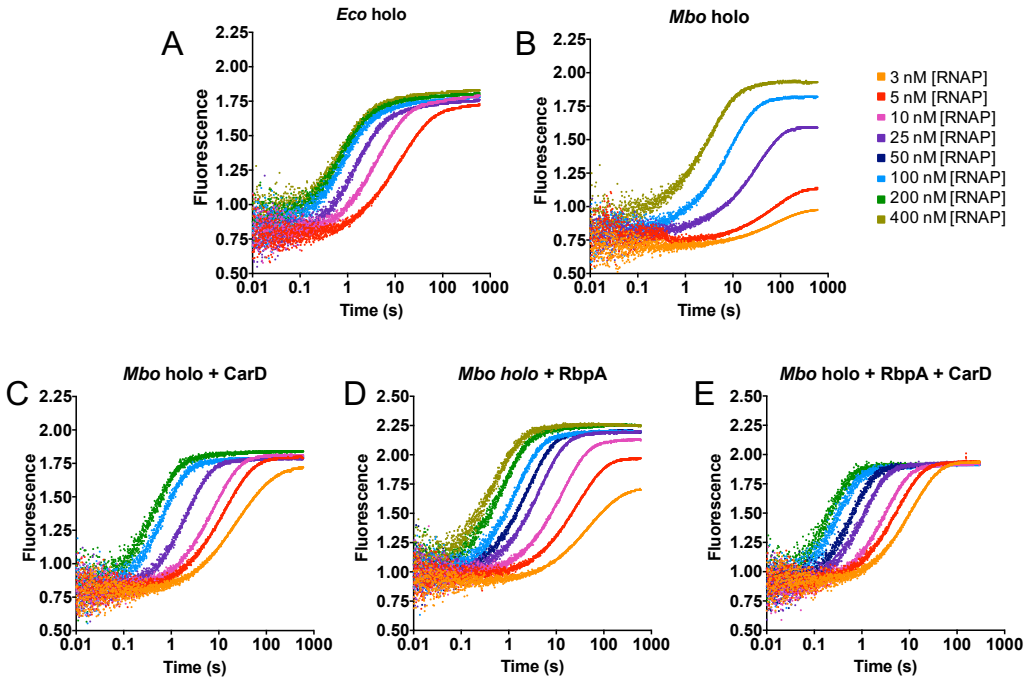


Figure 5.4: Traces of change in fluorescence intensity upon RPo formation with Cy3-AP3, plotted over time. For each condition reactions were performed with 1nM DNA for a series of RNAP concentrations (as indicated). When present, RbpA and CarD were at saturating conditions (5 µM and 10 µM, respectively). (A) *Eco* holo (B) *Mbo* holo (C) *Mbo* holo+CarD (D) *Mbo* holo+RbpA (E) holo+RbpA+CarD.

studies showing that compared to *Eco*, *Mbo* RNAP forms unstable and reversible open complexes (Davis et al., 2015).

Similar to *Mbo* holo alone, reactions with *Mbo* holo+RbpA also did not reach saturation at low or intermediate [RNAP] concentrations and took the same amount of time to level off (**Figure 5.4d**). CarD, however, clearly sped up *Mbo* RPo formation, taking ~100s for reactions containing 100 nM RNAP concentrations to level off (**Figure 5.2c**). Additionally, with CarD present, reactions reached saturation at lower RNAP concentration (**Figure 5.4c**), indicating that CarD stabilized RPo, which is also consistent with previous biochemical and structural studies (Davis et al., 2015; Bae et al., 2015). Strikingly, when both RbpA and CarD were added, reactions leveled off within 30s (similar to *Eco*; **Figure 5.4e**), consistent with the work presented in Chapter 4 indicating that the two factors work synergistically.

5.2 The kinetic data best fits a 3-step linear model for RPo formation

Rate constants for RPo formation are traditionally determined by measuring the observed rate of RPo formation (k_{obs}, provided by non-linear regressive fit of a time-course of RPo formation) at varying RNAP concentrations while keeping promoter DNA concentration constant. Plotting the calculated k_{obs} vs [RNAP] allows for direct calculation of individual rate or equilibrium constants (Ross and Gourse, 2009); however, with this method, a propagation of errors occurs. Rather than using this method to calculate rate constants for the Heyduk assay, we used KinTek Global Kinetic Explorer® software (Johnson et. al, 2008), which, for each sample, performs a global fit of the traces of all RNAP concentrations at once and uses the data from all of those fits to calculate

individual rate constants. Thus, this method is a more direct measurement of rates and lessons propagation of error.

We modeled the data on AP3 promoter to various mechanisms for RPo formation (**Figure 5.5, Appendix 2**), including a 1-step, 2-step, 3-step, 3-step branched, and 4-step mechanism. Based on previous work, we expected either a 2-step, 3-step (linear), or 4-step mechanism to apply. In our analysis of the data, we will make the assumption that the same kinetic model applies to each of the samples—i.e., *Mbo* and *Eco* holo form RPo using the same kinetic mechanism, and the presence of RbpA and CarD with *Mbo* holo does not change the mechanism itself but only the rate constants associated with that mechanism.

The KinTek program calculated rate constants, scale factors (the fluorescence contribution contributed by each intermediate species), and χ^2/DOF (DOF = Degrees of Freedom; a 'goodness of fit' measurement where a perfect fit =1.00) for each sample (**Appendix 2**). Based on the fitted kinetic constants, we were also able to calculate the theoretical RPo half-life ($t_{1/2}$) for each sample (Tsodikov et al., 1999) and compare those values to the experimental values for $t_{1/2}$ obtained in independent abortive initiation half-life assays (Chapter 4, **Figure 4.6**).

For each sample dataset, 1-step and 2-step models were rejected because 1) the curves obviously did not fit the data (**Figure 5.6**); 2) the χ^2/DOF was clearly greater than the ideal value of 1; and 3) in many cases, the calculated $t_{1/2}$ of the promoter complexes were inconsistent with the measured experimental values (**Appendix 2**). The 4-step model was also rejected; in every case, one of the fluorescence scale factors ('d', corresponding to the extra intermediate introduced in the 4-step pathway) always refined

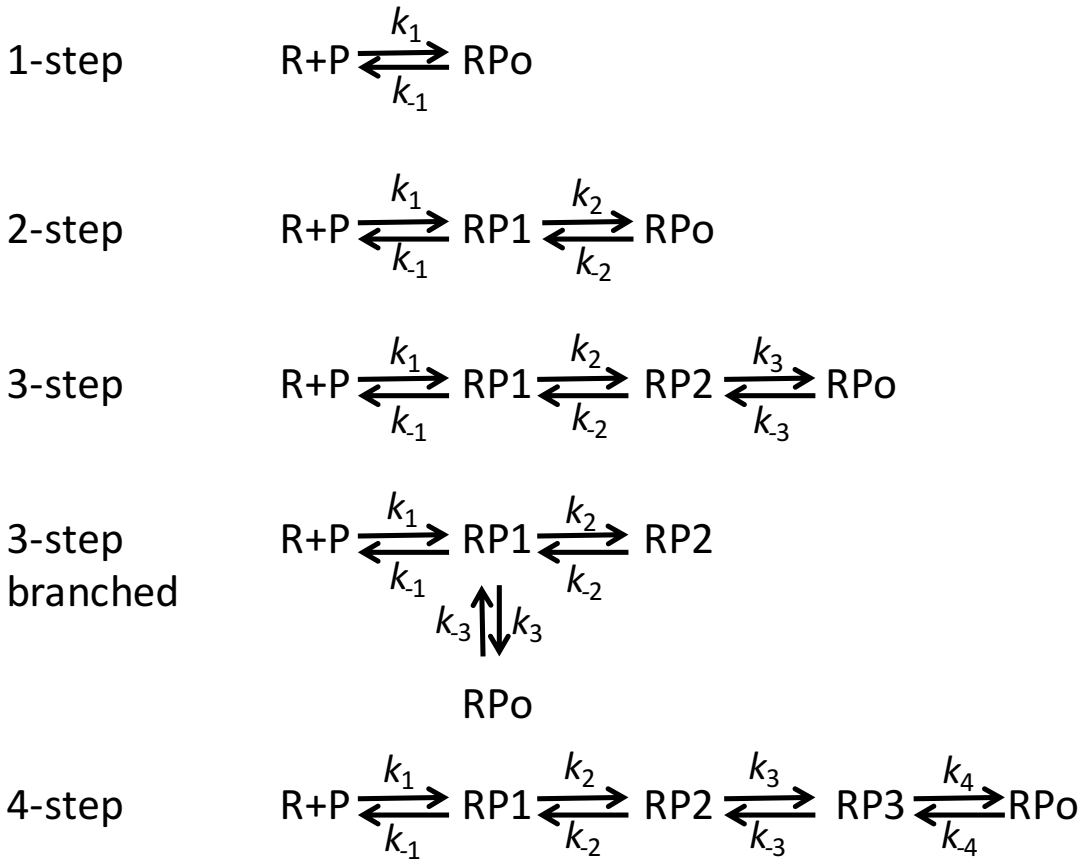


Figure 5.5: Models of RPo formation tested using KinTek Global Explorer (Johnson et al., 2009).

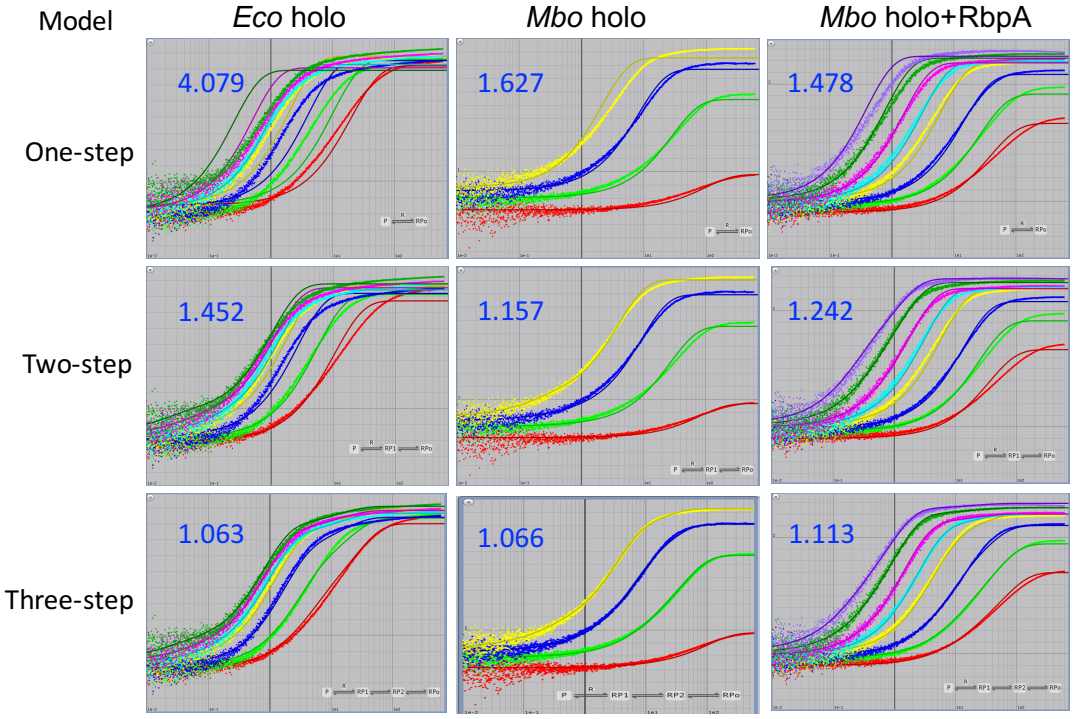


Figure 5.6: Mechanism for RPo formation on AP3 best fits a 3-step model. Representative KinTek Global Explorer Fits of Mbo holo and Eco holo. Blue values are χ^2/DOF (a goodness of fit measurement where a perfect fit = 1.00).

to a value near zero, indicating that adding this intermediate did not contribute to the fitting of the data and rendering the kinetic constants associated with this intermediate (k_3, k_{-3}, k_4, k_{-4}) meaningless. The 3-step branched model gave similar fits as the 3-step linear model, but in several cases the $t_{1/2}$ of the most populated product (RP2 or RP3) was inconsistent with the measured values (**Appendix 2**). Because we made the assumption that only one model applies to all samples, we rejected the 3-step branched model.

The 3-step (linear) model applies to each data set well: it visually fits the data; it gives χ^2/DOF values near 1 (ranging between 1.05 and 1.11); and for every sample, the calculated $t_{1/2}$ matches well with experimental values. Thus, for each condition (*Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+CarD, *Mbo* holo+RbpA+CarD, *Eco* holo), a concentration series was collected on a given day and fitted to the 3-step model. Data concentration series collected on separate days were fitted separately, yielding similar constants. The constants derived in this way were then averaged, giving rise to the average values and standard errors listed in **Table 5.2**.

The data fitting a 3-step mechanism indicate that the assay reports on three RNAP-promoter populations—the formation of two intermediate complexes (RP_1 and RP_2), and then, presumably, fully formed RPo (**Figure 5.7**). For the *Eco* holo and *Mbo* holo+RbpA+CarD (the two most rapidly occurring reactions), the forward rate for the initial bimolecular step (k_1) refines to the same value of 1.2 x 10^8 $M^{-1}s^{-1}$, possibly representing diffusion-limited binding. If this were the case, this binding step would not involve any conformational changes in the RNAP or DNA; thus RP_1 may represent RPc. Another observation is that k_3 is remarkably similar across all of the samples (the standard deviation of the average value across all the samples is only 26 %; **Table 5.2**).

Table 5.2: kinetic values for Heyduk assay on Cy3-AP3 before constraining k_3 and k_{-3}

	RNAP				
	M. bovis holo				Eco
		+RbpA	+ CarD	+CarD+RbpA	
k_1 ($M^{-1}s^{-1}$)	(9.6 ± 0.2) x 10^6	(1.7 ± 0.3) x 10^7	(4.2 ± 0.6) x 10^7	(1.2 ± 0.0) x 10^8	(1.2 ± 0.1) x 10^8
k_{-1} (s^{-1})	3.2 ± 1.2	1.8 ± 0.6	3.2 ± 0.9	34 ± 12	3.8 ± 0.6
K_1 (M^{-1})	(3.0 ± 1.1) x 10^6	(9.4 ± 3.6) x 10^6	(1.3 ± 0.4) x 10^7	> (3.6 ± 1.2) x 10^6	> (3.2 ± 0.6) x 10^7
k_2 (s^{-1})	0.40 ± 0.04	1.6 ± 0.3	3.3 ± 0.8	10 ± 1	1.2 ± 0.03
k_{-2} (s^{-1})	0.074 ± 0.005	0.086 ± 0.012	0.078 ± 0.007	0.024 ± 0.008	0.082 ± 0.006
K_2	5.4 ± 0.6	18 ± 3	43 ± 11	429 ± 150	15 ± 1
k_3 (s^{-1})	0.067	0.067	0.067	0.067	0.067
k_{-3} (s^{-1})	0.013 ± 0.001	0.016 ± 0.001	(2.3 ± 0.1) x 10^{-3}	(1.6 ± 0.4) x 10^{-3}	1.5 x 10^{-7}
K_3	5.4	4.6 ± 0.3	32 ± 1	46 ± 11	4.5 x 10^5
$K_1K_2K_3$	(8.7 ± 3.3) x 10^7	(7.9 ± 3.2) x 10^8	(1.8 ± 0.7) x 10^{10}	(7.0 ± 3.8) x 10^{10}	2.2 x 10^{14}
k_a ($M^{-1}s^{-1}$)	(9.3 ± 4.2) x 10^6	(1.7 ± 0.9) x 10^7	(4.2 ± 2.3) x 10^7	(1.2 ± 0.8) x 10^8	1.3 x 10^8
k_i (s^{-1})	0.057 ± 0.002	0.080 ± 0.003	0.072 ± 0.003	0.074 ± 0.003	0.060 ± 0.001
k_d	(6.0 ± 0.4) x 10^{-3}	(7.5 ± 1.8) x 10^{-3}	(1.1 ± 0.5) x 10^{-3}	(3.8 ± 0.8) x 10^{-4}	8.2 x 10^{-8}
$t_{1/2}$ (min)	1.9 ± 0.1	1.5 ± 0.4	11 ± 06	30 ± 6	1.4 x 10^5
$t_{1/2}^{exp}$ (min)	~2	~1.5	~10	~30	>> 60

Light Green: 5-10 fold faster than *Mbo* holo
Dark Green: > 10 fold faster than *Mbo* holo

Light Red: 5-10 fold slower than *Mbo* holo
Dark Red: > 10 fold slower than *Mbo* holo

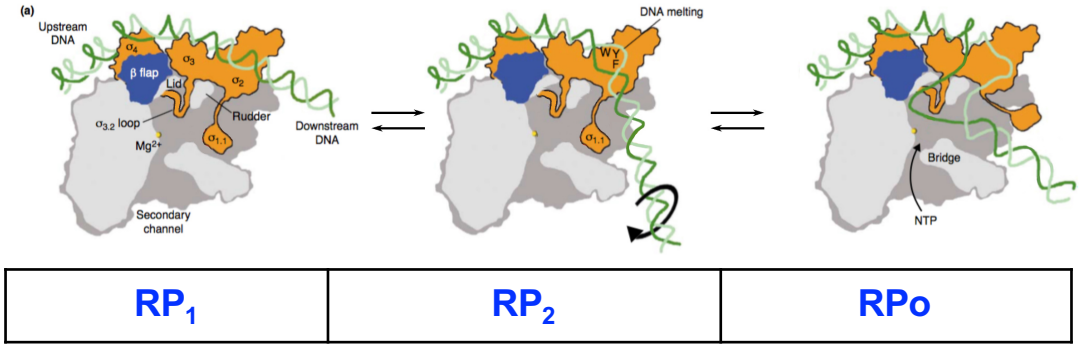


Figure 5.7 Intermediates of RPo formation on AP3 are consistent with those defined on *Eco* promoters. Hypothetical models of intermediates of RPo formation identified in Heyduk assay on AP3.

This rate is apparently not influenced by the holoenzyme (*Mbo* or *Eco*) or by the presence of RbpA and/or CarD, suggesting that k_3 may be controlled by the properties of the DNA. In previous kinetic studies on various *Eco* promoters, the rate-limiting step of RPo formation is the step involving the full melting of the transcription bubble and placement of the template strand into the channel to form RPo. That we found k_3 to be the limiting step in our 3-step fitted model of RPo formation on AP3 further indicates that this step may reflect melting of the DNA to form the full transcription bubble.

Although the fluorescence scale factors (a,b,c, and d) are not constrained during fitting, and thus can be any value, we were encouraged that they all refine to relatively consistent values across all samples (**Tables 5.2**), as the contribution of each state toward the fluorescence signal should not be influenced by whether or not factors are present unless the factor directly alters the environment of the fluorophore. Our structural analysis (Chapter 3) leads us to expect that this is not the case. Both CarD and RbpA bind to the RNAP and interact with the promoter DNA at the upstream edge, or just upstream of the -10 element, which is far (~45 Å for linear DNA) from the Cy3-attachment site at +2. Thus, these values (a, b, c, d) may be physically meaningful. The value of 'a', the scale factor for the free promoter DNA, refines to the lowest value (~ 0.27 ± 0.02). The fluorophore attached to the free promoter would be the most solvent exposed and would be expected to have the lowest fluorescence intensity; the scale factor for RP1 ('b') refines to an intermediate value (~ 0.45 ± 0.06), which suggests that RP1 represents an intermediate state where the fluorophore environment has become more proteinaceous but is still relatively solvent exposed; the scale factors for RP2 ('c') and RPo ('d') refine

to the highest values (~1.2 ± 0.04, ~1.2 ± 0.02, respectively) suggesting that in these two states, the fluorophore is shielded from solvent (*i.e.*, inside the RNAP active site channel).

We performed additional validation of our data using the KinTek FitSpace software (Johnson et al., 2008), which determines if the rate constants calculated by the program are constrained. We noticed that some of the fitted rate constants were not well constrained; in other words, other values of some of the constants for some of the samples could yield fits with similar statistics, indicating that the fits were not unique. To constrain the modeling further and obtain more valid constants, we performed the following procedures:

i) We noticed that the calculated RPo $t_{1/2}$ was most sensitive to the value of k_{-3}. Therefore, for each dataset, we adjusted the value of k_{-3} until the calculated RPo $t_{1/2}$ from the 3-step model was equal to the experimentally determined value. We then fixed k_{-3} and refined the other constants against the data. We iteratively adjusted k_{-3} and refined the other values so that the final calculated RPo $t_{1/2}$ was equal to the experimentally determined value, and the refinement of the other constants was stable (the constants didn't change anymore). Since the initial (unrestrained) refinement resulted in calculated RPo $t_{1/2}$ values that were not far from the experimental values, this did not require dramatic changes of the constants, but resulted in a final refinement where k_{-3} was fixed.

ii) Since we noticed k_3 was very similar across all the samples (**Table 5.2**), we fixed k_3 to a constant value that worked well for all of the datasets (0.067 s^{-1}; **Table 5.3**). The recalculated rate constants closely fit the original values (**Tables 5.2, 5.3**), and the Fitspace validation showed that the calculated parameters were constrained by the data (**Figure 5.8**) with the exception of k_{-1} and k_{-2} for holo+RbpA+CarD (**Figure 5.8**). For

Table 5.3: kinetic values for Heyduk assay on Cy3-AP3 after constraining k_3 and k_{-3}

	RNAP				
	M. bovis holo				Eco
		+RbpA	+ CarD	+CarD+RbpA	
k_1 ($M^{-1}s^{-1}$)	(9.6 ± 0.2) x 10^6	(1.7 ± 0.3) x 10^7	(4.2 ± 0.6) x 10^7	(1.2 ± 0.0) x 10^8	(1.2 ± 0.1) x 10^8
k_{-1} (s^{-1})	3.2 ± 1.2	1.8 ± 0.6	3.2 ± 0.9	34 ± 12	3.8 ± 0.6
K_1 (M^{-1})	(3.0 ± 1.1) x 10^6	(9.4 ± 3.6) x 10^6	(1.3 ± 0.4) x 10^7	> (3.6 ± 1.2) x 10^6	> (3.2 ± 0.6) x 10^7
k_2 (s^{-1})	0.40 ± 0.04	1.6 ± 0.3	3.3 ± 0.8	10 ± 1	1.2 ± 0.03
k_{-2} (s^{-1})	0.074 ± 0.005	0.086 ± 0.012	0.078 ± 0.007	0.024 ± 0.008	0.082 ± 0.006
K_2	5.4 ± 0.6	18 ± 3	43 ± 11	429 ± 150	15 ± 1
k_3 (s^{-1})	0.067	0.067	0.067	0.067	0.067
k_{-3} (s^{-1})	0.013 ± 0.001	0.016 ± 0.001	(2.3 ± 0.1) x 10^{-3}	(1.6 ± 0.4) x 10^{-3}	1.5 x 10^{-7}
K_3	5.4	4.6 ± 0.3	32 ± 1	46 ± 11	4.5 x 10^5
$K_1K_2K_3$	(8.7 ± 3.3) x 10^7	(7.9 ± 3.2) x 10^8	(1.8 ± 0.7) x 10^{10}	(7.0 ± 3.8) x 10^{10}	2.2 x 10^{14}
k_a ($M^{-1}s^{-1}$)	(9.3 ± 4.2) x 10^6	(1.7 ± 0.9) x 10^7	(4.2 ± 2.3) x 10^7	(1.2 ± 0.8) x 10^8	1.3 x 10^8
k_i (s^{-1})	0.057 ± 0.002	0.080 ± 0.003	0.072 ± 0.003	0.074 ± 0.003	0.060 ± 0.001
k_d[a]	(6.0 ± 0.4) x 10^{-3}	(7.5 ± 1.8) x 10^{-3}	(1.1 ± 0.5) x 10^{-3}	(3.8 ± 0.8) x 10^{-4}	8.2 x 10^{-8}
$t_{1/2}$ (min)	1.9 ± 0.1	1.5 ± 0.4	11 ± 06	30 ± 6	1.4 x 10^5
$t_{1/2}^{exp}$ (min)	~2	~1.5	~10	~30	>> 60

Light Green: 5-10 fold faster than *Mbo* holo
Dark Green: > 10 fold faster than *Mbo* holo

Light Red: 5-10 fold slower than *Mbo* holo
Dark Red: > 10 fold slower than *Mbo* holo

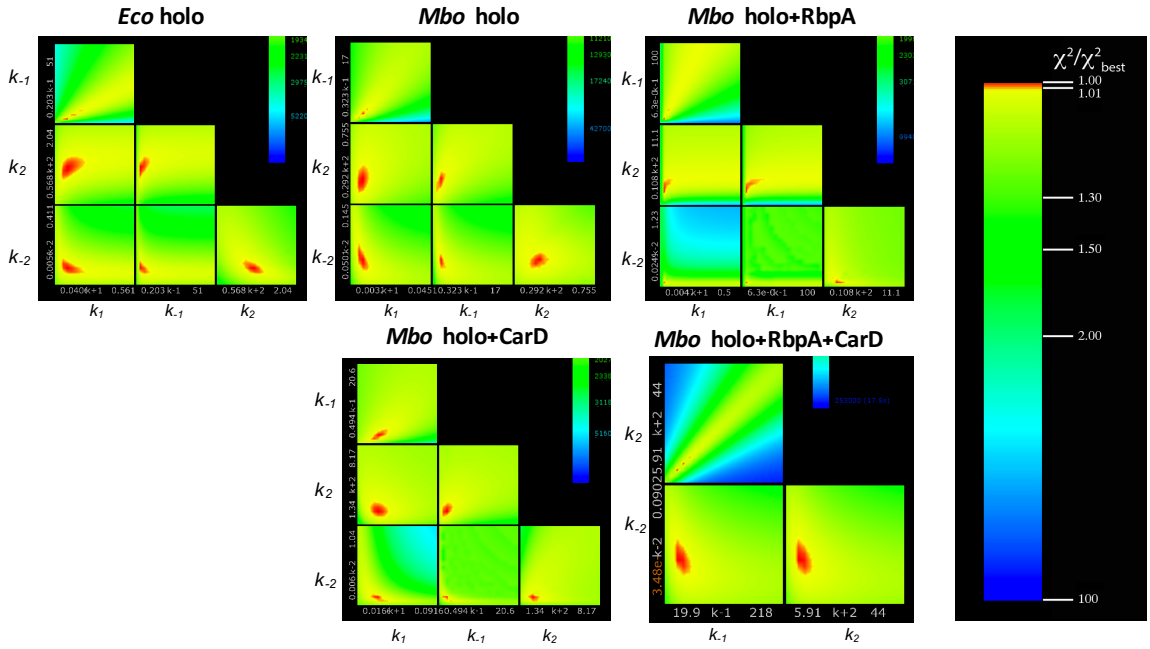


Figure 5.8: FitSpace validations of calculated kinetic rate constants for AP3. FitSpace was performed for the 3-step mechanism of RPo formation by *Eco* holo, *Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+CarD, *Mbo* holo+RbpA+CarD. FitSpace validation varies two rate constants at a time for a large range of values. For each pair of values, the program calculates χ^2 (colored according to right panel). Red indicates low χ^2; therefore, the smaller the red region, the more constrained the calculated rate constant. Above panels show validation that was performed after constraining k_3 and k_{-3}. For *Mbo* holo+RbpA+CarD, k_1, was also constrained (which allowed for good FitSpace validation, but did not change the calculated rate constants).

these data, the refined k_1 was reasonably restrained ((1.2 ± 0.6) x 10^8 $M^{-1}s^{-1}$), so we performed another refinement with k_1 locked at this value, resulting in good constraints for the freely refined constants k_{-1}, k_2, and k_{-2} (**Figure 5.8**). Additional data for *Mbo* holo+RbpA+CarD need to be collected over a wider range of [RNAP] to provide additional constraints for the fitting.

In summary, our Heyduk assay data on the AP3 promoter convincingly fit a three-step model and is consistent with previously defined intermediate states characterized in a three-step mechanism for RPo formation. Also of note is that the calculated rate constants for this model were similar to those observed in the three-step mechanism for RPo formation with *Eco* RNAP on T7A1 (Sclavi et al., 2006).

Although the kinetic model had too many free-floating parameters to rigorously and uniquely fit the datasets, several observations led us to conclude that these initial estimates of the kinetic parameters (**Table 5.2**) were physically meaningful:

1. Most importantly, the freely-refined parameters for each dataset allowed calculation of the expected $t_{1/2}$ for the final RPo (Tsodikov et al., 1999). These calculated $t_{1/2}$ values matched well with experimental values (**Table 5.2**) determined completely independently (Chapter 4).

2. The k_3 across all of the samples refined to very similar values (the standard deviation of the measurements over all of the samples was only 26% of the average value; **Table 5.2**).

3. The fitted parameters for matching datasets collected independently on separate days (often over different [RNAP] ranges) were closely matched (SEMs in **Table 5.2**).

4. The fluorescence scale factors (a, b, c, d) refined to similar values across all of the samples (**Table 5.2**).

This emboldened us to impose additional constraints on the kinetic model with the goal of reducing the number of free-floating parameters: i) The refined parameters were forced to yield the experimentally determined RPo $t_{1/2}$, and ii) k_3 was locked at an 'average' value. These two constraints allowed rigorous, well constrained refinement of the remaining kinetic parameters (k_1, k_{-1}, k_2, k_{-2}) for the *Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+CarD, and *Eco* holo data (**Figure 5.8**). An additional constraint (locking k_1) was required for the *Mbo* holo+RbpA+CarD dataset (**Figure 5.8**).

5.3 RbpA and CarD cooperatively drive formation of RP2:

The activation energies required for each kinetic step were calculated by the Kintek Global Kinetic Explorer® (**Figure 5.9**) based on the fitted rate constant values using the equation:

$$\Delta G_i^{\dagger} = RT \ln\left(\frac{k_B T / h}{k_i}\right)$$

Where $\Delta G^{\ddagger}_i$ = activation energy, $h = 6.62e^{-34}\ J{\bullet}s$, $k_B = 1.38e^{-34}\ J/K$, and k_i = rate constant variable. Note that for all of the samples, the highest energy barrier (and thus the rate limiting step) in the forward direction is for RP2 → RPo, and the same step in the reverse direction is also rate limiting (RPo → RP2).

In the absence of any factors, the most striking difference between *Mbo* and *Eco* is the energy barrier for the conversion of RPo→RP2 (bubble collapse), which is significantly larger in *Eco* than in *Mbo* because of a much smaller k_{-3} value in *Eco*

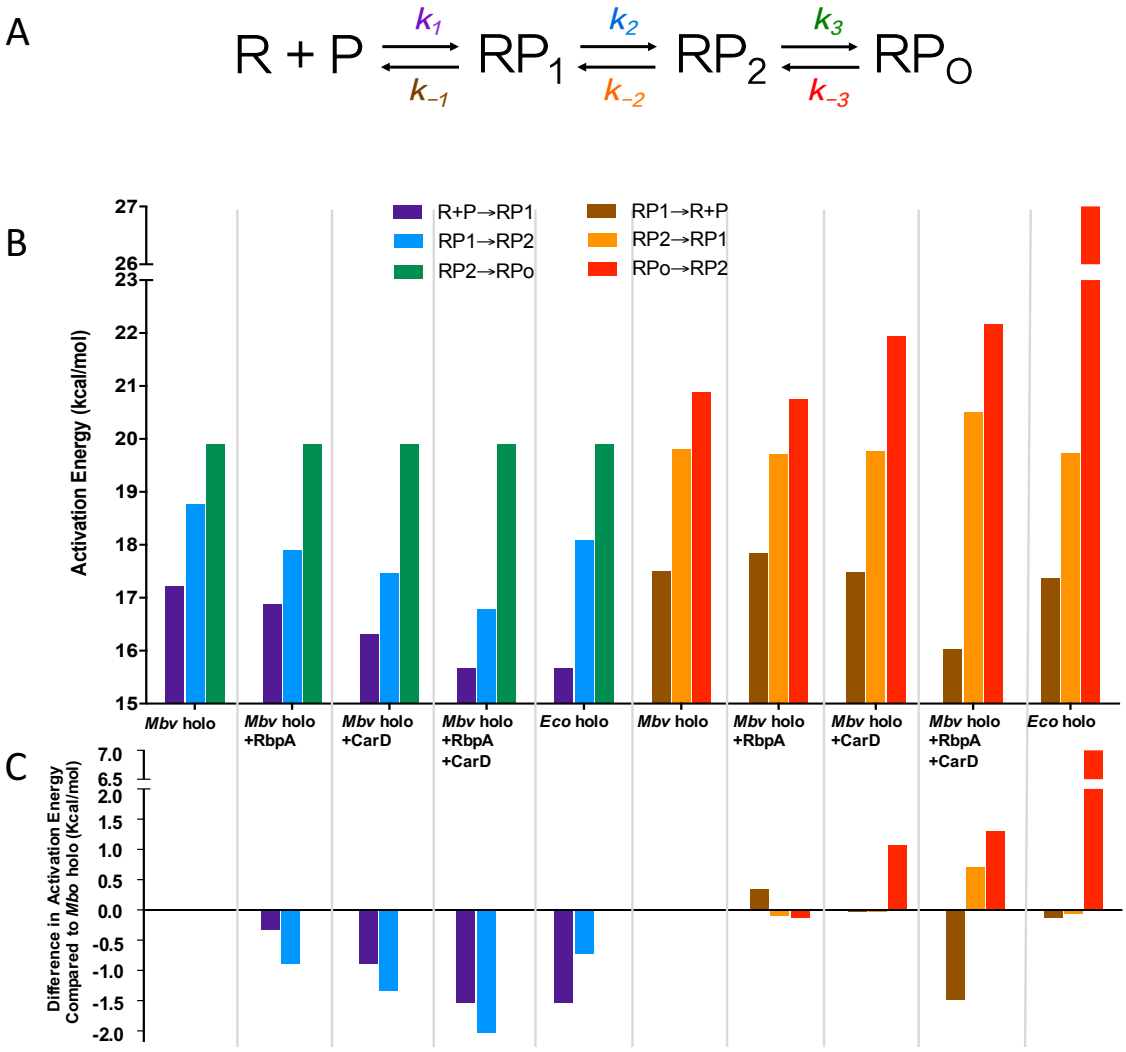


Figure 5.9: The effect of RbpA and CarD on energy required for steps in RPo formation on AP3. (A) Three step mechanism for RPo formation on AP3 as measured by the Heyduk assay. (B) Activation energy for individual kinetic steps in RPo formation. (C) Difference in activation energy compared to *Mbo* holo.

compared to *Mbo* (1.5 x10^{-7} and 0.013, respectively; **Table 5.3**). This difference accounts for the essentially irreversible open complex formed by *Eco* and the relatively unstable RPo formed by *Mbo* observed in *in vitro* half-life assays (Davis et al., 2015). The addition of CarD to *Mbo* significantly stabilizes RPo by increasing the activation energy required for bubble collapse (RPo→RP2) by ~1 kcal/mol (**Figures 5.9, 5.10**). This effect is consistent with structural characterization of *Thermus* CarD bound to RPo, which shows that CarD physically prevents the bubble collapse through its wedging mechanism (Bae et al., 2015). CarD also increases the rate of two forward steps (k_1 and k_2 **Table 5.3**), first by modestly decreasing the energy barrier for formation of RP1 and then more significantly driving conversion into RP2 by decreasing the energy barrier required for RP1→RP2 by 1.3 kcal/mol (**Figures 5.9, 5.10**).

When RbpA is added alone, its effect on the AP3 promoter is mild (similar to its activation in transcription assays), as it modestly lowers the energy barrier for RP2 formation (**Figures 5.9, 5.10**), increasing the equilibrium constant K_2 about 3 fold (**Table 5.3**). However, when added together, RbpA and CarD have a cooperative effect on K_2 (increasing it ~80 fold) (**Table 5.3**), lowering the energy barrier for the formation of RP2 by 2.0 kcal/mol (**Figures 5.9, 5.10**), and thus causing a substantial increase in the population of the second intermediate (**Figure 5.11**). The larger population of RP2 would push the reaction toward the formation of RPo, which would then be stabilized by CarD, thus explaining the striking synergy between RbpA and CarD in half-life assays (Chapter 4).

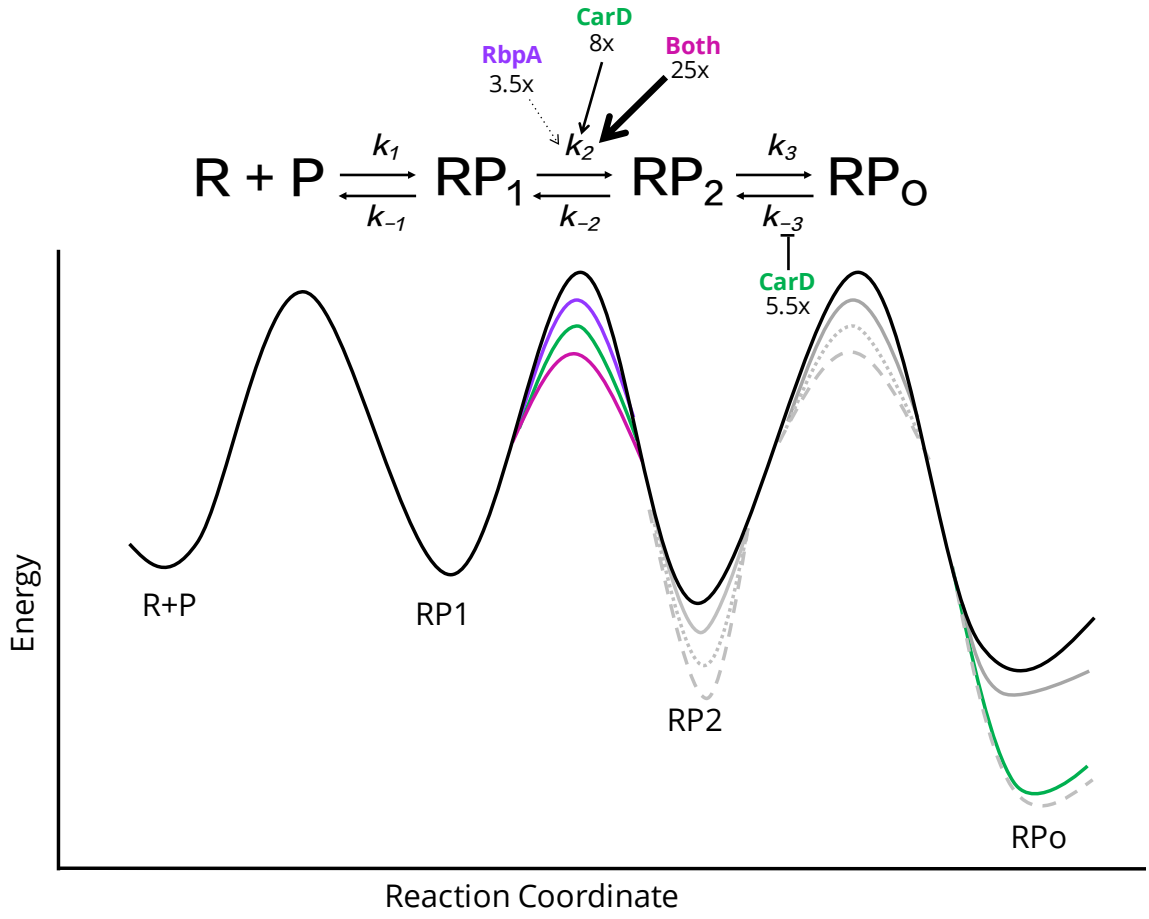


Figure 5.10 Top: Reaction scheme showing major effects of RbpA and CarD. Bottom: Illustrative reaction coordinate diagram depicting change in activation energy. RbpA (purple and solid gray lines) decreases activation energy for RP1→ RP2 (purple line). CarD (green and dotted gray lines) decreases activation energy for RP1→ RP2 and stabilizes RPo (green line). RbpA and CarD together (magenta and dashed gray line) decrease activation energy for RP1→ RP2 (magenta line) synergistically.

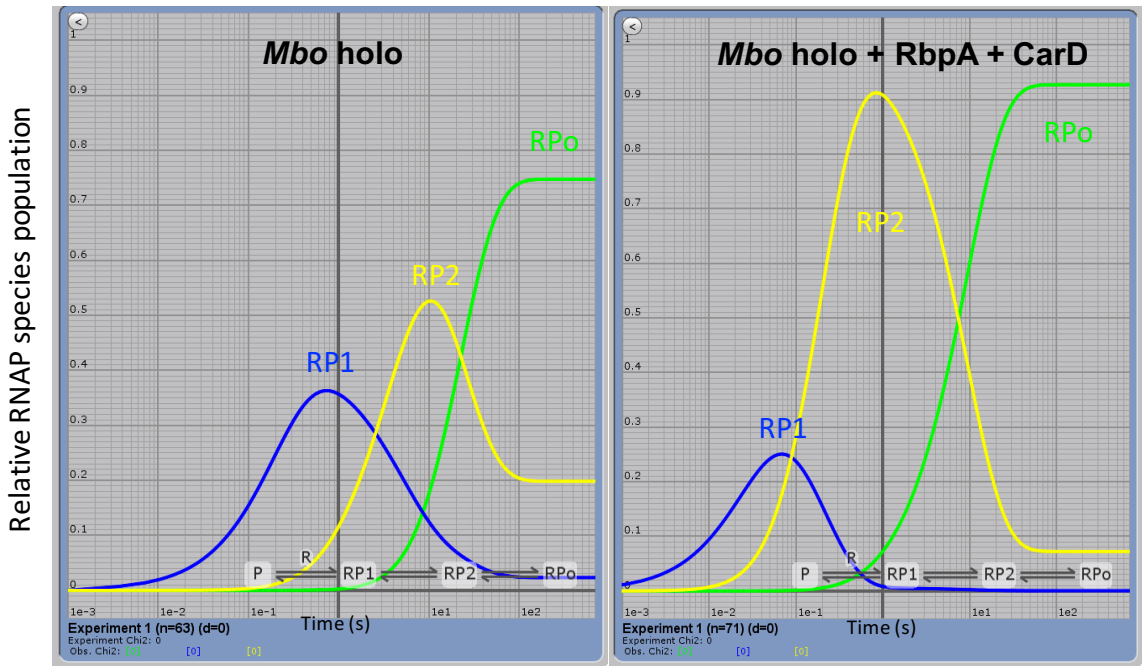


Figure 5.11 RbpA and CarD increase the population of the RP2 intermediate. KinTek Global Explorer simulation of RNAP populations over time based on a 3-step mechanism of RPo formation and kinetic constants calculated from AP3 data. The x-axis is time (log) and the Y axis is relative population. When added together, RbpA and CarD significantly speed up the conversion of RP1 to RP2 causing a large increase in RP2, pushing equilibrium toward conversion into RPo. After 100 s, ~75% of the of the RNAP species were RPo, while with factors >90% of the RNAP species were RPo.

Interestingly, with RbpA and CarD present, the rate of formation of RP2 (k_2) by *Mbo* holo is much larger than that of *Eco* holo (34 ± 12 vs 3.8 ± 0.6 s^{-1}, respectively; **Table 5.3**), indicating that the strong activity of *Eco* holo can be primarily attributed to its very stable RPo.

5.4 The RbpA-$^{SID-BL}$ increases K_2 to a larger extent than WT

Because full-length RbpA and RbpA truncations activate more on the VapB promoter than on AP3, we decided to test their effect on a Cy3-labeled VapB promoter (bases -61+53; **Figure 5.12**) in hopes of better understanding the function of the RbpA structural elements. Experiments were carried out similar to those with AP3 (described in Section 5.1 and 5.2), with some adjustments in DNA and RNAP concentration (**Table 5.4**). The data were fit to a 3-step model, as we assumed that the intermediates detected on VapB would be the same on AP3; however, we are currently in the process of validating this model by fitting other models (2-step, 4-step, etc.) and also running fit space. Again for the fitted parameters, k_3 refined to very similar values across all the samples (0.021 ± 0.008 std. deviation), and the fluorescence scale factors refined to consistent values (mean ± std. deviation; a = 0.23 ± 0.02; b = 0.56 ± 0.11; c = 1.7 ± 0.5; d = 1.3 ± 0.1) that are also similar to the values for the AP3 data, suggesting that the intermediates giving rise to these fluorescence signals could be similar (**Table 5.5**). We didn't see a significant change in the fitted rate constants for RPo formation on AP3 after adding constraints; therefore, we don't expect the fitted constants for the Heyduk assay on VapB to change significantly after refinement.

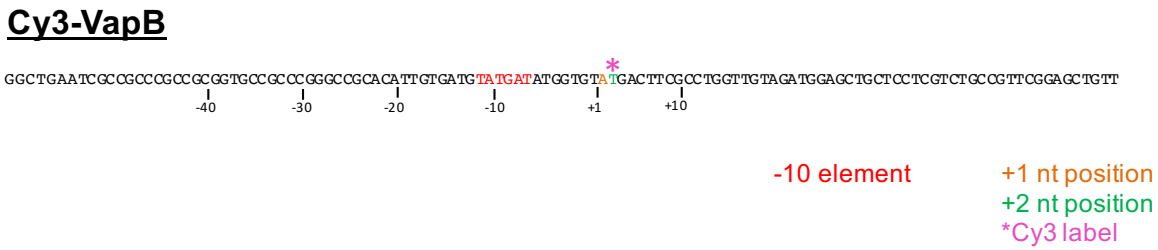


Figure 5.12: Cy3-Vap DNA used for Heyduk Assay. (A) Non-template strand sequence of Cy3-AP3 DNA. Elements are colored as indicated. Cy3 label (pink) is attached to the +2 thymine base (green).

Table 5.4: Concentrations of protein and DNA for Heyduk Assay Experiment with Cy3-VapB

Sample	[Cy3-VapB]	[RNAP] range	[RbpA Derivatives]
Mbo holo	2 nM	10-500 nM	-
Mbo holo +RbpA	2 nM	10-500 nM	5 µM
Mbo + $RbpA^{SID-BL}$	2 nM	10-500 nM	5 µM
Mbo holo + $RbpA^{\Delta NTD}$	2 nM	10-500 nM	5 µM
Mbo holo + $RbpA^{R79A}$	2 nM	10-500 nM	5 µM

Table 5.5: Kinetic values for Heyduk assay on Cy3-VapB

	RNAP				
	M. bovis holo				
		+RbpA	+RbpA-SID-BL	+RbpA-R79	+RbpA-ΔNTD
k_1 ($M^{-1}s^{-1}$)	(1.9 ± 0.3) x 10^7	(1.9 ± 0.2) x 10^7	(1.6 ± 0.1) x 10^7	(1.4 ± 0.2) x 10^7	(1.3 ± 0.1) x 10^7
k_{-1} (s^{-1})	1.6 ± 0.2	2.7 ± 0.4	3.1 ± 0.4	0.82 ± 0.06	3.8 ± 1.5
K_1 (M^{-1})	(1.2 ± 0.2) x 10^7	(7.2 ± 1.2) x 10^6	(5.3 ± 0.7) x 10^6	(1.7 ± 0.3) x 10^7	(3.3 ± 1.3) x 10^6
k_2 (s^{-1})	0.031 ± 0.002	0.17 ± 0.02	0.39 ± 0.04	0.013 ± 0.006	0.19 ± 0.04
k_{-2} (s^{-1})	0.040 ± 0.017	0.063 ± 0.022	0.038 ± 0.001	0.034 ± 0.020	0.039 ± 0.007
K_2	0.77 ± 0.33	2.7 ± 1.0	10 ± 1	0.39 ± 0.29	4.9 ± 1.3
k_3 (s^{-1})	0.013 ± 0.001	0.033 ± 0.017	0.015 ± 0.003	0.023 ± 0.002	0.022 ± 0.006
k_{-3} (s^{-1})	(3.0 ± 0.9) x 10^{-3}	(3.6 ± 0.1) x 10^{-3}	(3.9 ± 0.7) x 10^{-3}	(7.9 ± 1.4) x 10^{-3}	0.0032 ± 0.0003
K_3	4.2 ± 1.2	9.2 ± 4.7	3.8 ± 0.9	2.9 ± 0.5	6.8 ± 1.8
$K_1K_2K_3$	(3.9 ± 2.1) x 10^7	(1.8 ± 1.2) x 10^8	(2.0 ± 0.6) x 10^8	(1.9 ± 1.5) x 10^7	(1.1 ± 0.6) x 10^8
k_d	(2.1 ± 1.3) x 10^{-3}	2.3 x 10^{-3}	(2.6 ± 0.6) x 10^{-3}	(4.1 ± 3.8) x 10^{-3}	(1.9 ± 0.8) x 10^{-3}
$t_{1/2}$ (min)	5.5 ± 3.3	5.1	4.5 ± 1.1	2.8 ± 2.6	6.0 ± 2.4
a	0.26	0.22	0.22	0.25	0.23
b	0.47	0.62	0.68	0.41	0.62
c	1.5	1.6	1.3	2.6	1.4
d	1.4	1.4	1.3	1.1	1.3
bkg	0.91	1.1	0.98	0.97	1.0

Values from *Mbo* holo on VapB are compared to *Mbo* holo on AP3. All other conditions are compared to *Mbo* holo on VapB.

Light Green: 5-10 fold faster
Dark Green: > 10 fold faster
Light Red: 5-10 fold slower
Dark Red: > 10 fold slower

Upon performing the Heyduk assay on VapB, we noticed that compared to AP3, the reactions took significantly longer to level off (>45 min for the lowest concentration) (**Figure 5.13**). Interestingly, the fitted rate constants for the VapB data indicate that unlike AP3, in the absence of factors (*Mbo* holo alone), both the second (RP1 → RP2) and third (RP2 → RP3) steps represent significant energy barriers in the forward direction. In the reverse direction, the step RPo → RP2 is the clear rate limiting step. RbpA modulates the rate of RPo formation by lowering the energy barrier for the second step (RP1 → RP2) and has little effect on any of the other steps (forward or reverse).

Consistent with the results on AP3, RbpA increases the formation of the second intermediate by decreasing the activation energy required for RP1→RP2 (**Figure 5.14**), increasing the K_2 equilibrium constant 3.5 fold (**Table 5.5**). Consistent with its greater activation in transcription assays, the RbpA SID-BL decreases this energy barrier to a larger extent than WT (**Figure 5.14**), increasing K_2 equilibrium constant almost 13 fold (**Table 5.5**). Further, compared to *Mbo* holo alone, RbpA R79A mutant *increases* the energy required for RP2 formation by 0.6 kcal/mol, while also *decreasing* the activation energy of the bubble collapsing by 0.6 kcal/mol (**Figure 5.14**), accounting for it decreasing the $t_{1/2}$ in half-life assays (**Figure 4.6, Table 5.5**) and inhibition in transcription assays (**Figure 4.4**). These findings provide further support that the SID-BL is responsible for RbpA's activating function--driving the forward production of the RP2 intermediate. It also confirms that the RbpA-CD has a role independent of transcription initiation, and, in fact, appears to destabilize RPo formation.

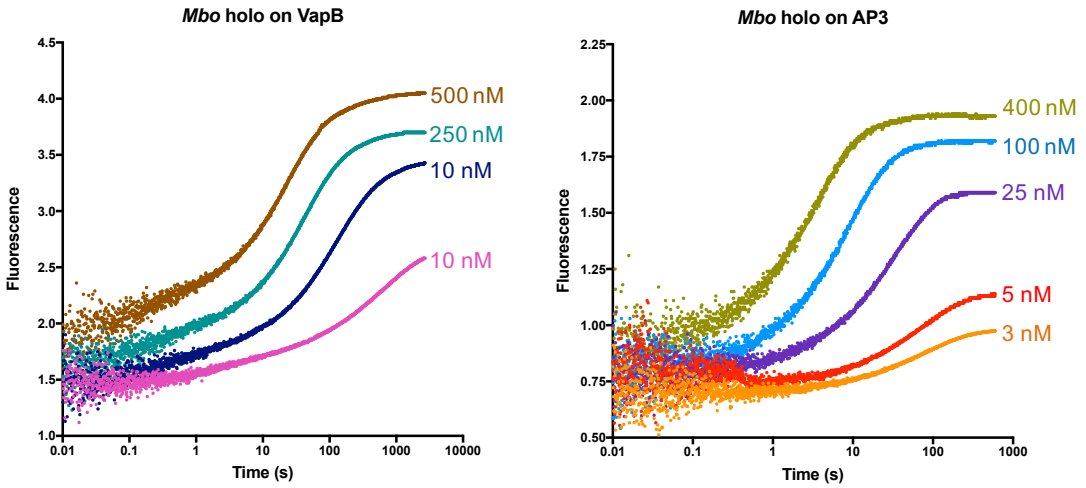


Figure 5.13: Traces of RPo formation on by *Mbo* holo on VapB and AP3. Reactions with VapB (left) take significantly longer to level off than on AP3 (right).

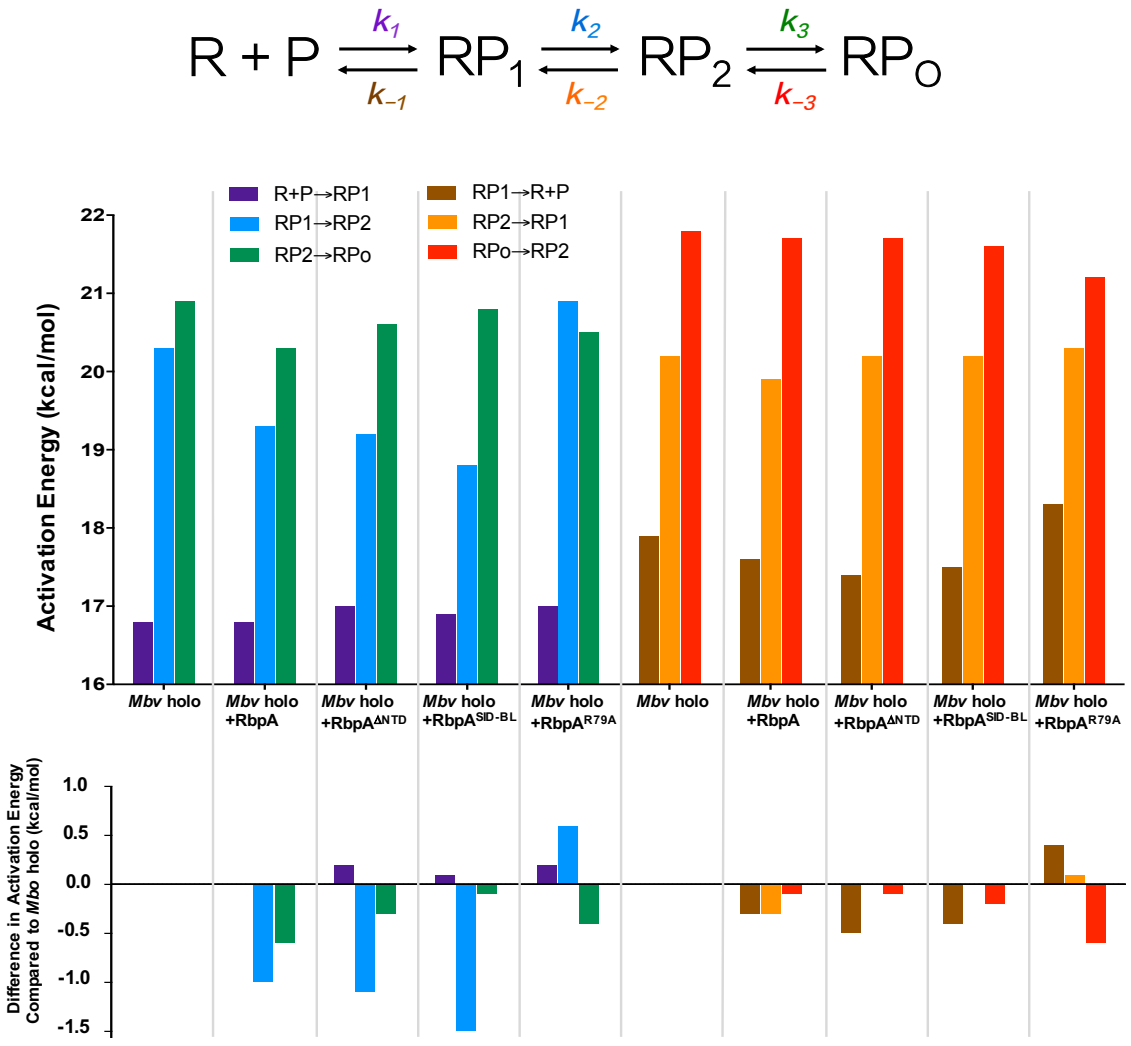


Figure 5.14: The effect of RbpA derivatives on energy required for steps in RPo formation on VapB. (A) Three step mechanism for RPo formation on vapb as measured by the Heyduk assay. (B) Activation energy for individual kinetic steps in RPo formation. (C) Difference in activation energy compared to *Mbo* holo.

5.5 Discussion:

A recent study used the Heyduk fluorescence assay to show that RbpA and CarD have a cooperative effect on RPo formation (Rammohan et al., 2016); however, a mechanism for the combined effect was not reported. Our data and use of the Kintek Global Explorer software allowed us to estimate the individual rates of a three-step mechanism for RPo formation, determine how RbpA and CarD cooperatively drive formation of intermediates along this pathway, and propose a mechanism for transcription activation by RbpA.

Individually, RbpA and CarD both drive formation of the RP2 intermediate, while CarD also slows bubble collapse (RPo → RP2). When added together, RbpA and CarD have a cooperative effect on lowering the energy required for formation of RP2. Their effect causes a synergistic increase in K_2 and subsequently a large increase in the population of the RP2 (**Tables 5.3, 5.5**). This effect not only provides us with a model accounting for their synergy in *in vitro* abortive initiation assays, but it also explains how the addition of RbpA enhances CarD's effect on RPo half-life: the larger population of RP2 helps prevent bubble collapse by pushing equilibrium toward the formation of RPo, and once RPo is formed, its dissociation is significantly slowed by CarD. That CarD is essential for bubble stability suggests why RbpA is unable to extend half-life by itself.

We expected that because VapB has no -35 or extended -10, k_1 (the forward rate of RP1 formation) would be much slower than that of AP3. However, the major difference between the two promoters is the value of k_2 (the forward rate of RP2 formation), which is more than 10 fold faster on AP3 (**Tables 5.3, 5.5**). Thus, sequence-specific interactions made by σ_4 with the -35 element do not significantly drive formation

of the initial RNAP/DNA complex (RP1, presumably RPc), but instead aids RP2 formation. It is possible that anchoring the -35 region aids in the -10 element melting and DNA bending required for the formation of RP2.

Based on these findings, we propose a similar anchoring role for the RbpA SID-BL in facilitating formation of RP2: the R79 contact with DNA just upstream of the bubble (**Figure 3.13**) may stabilize bending and subsequent melting of promoter DNA. An anchoring role of RbpA could also explain why the combined effect of RbpA and CarD on K_2 is synergistic in nature (more than additive), suggesting that the presence of one activator enhances the effect of the other. We know from structural studies that CarD is incompatible with a closed complex (Bae et al., 2015), so if RbpA facilitates the formation of RP2 (which is coupled to -10 element melting), it would, in turn, allow CarD to better bind the initiation complex, thus providing further explanation as to how RbpA enhances CarD's effect on RPo half-life.

Chapter acknowledgements:

Kinetic analysis using KinTek Global Explorer was performed with the assistance of Dr. Seth Darst.

Chapter 6:

A structural comparison of *Msm*, *Eco,* and *Taq* RNAP initiation complexes

Although the core RNAP subunits are highly conserved in sequence and structure between the three kingdoms, and even more so within bacteria, the larger subunits (β and β') in bacteria contain inserts that are lineage specific (lineage specific insertions or LSIs) to certain clades (Iyer, et al. 2004; Lane et al., 2010a; 2010b). These LSIs have been proposed to modulate the activity of the enzyme, either directly or by interacting with cis-acting regulators (Artsimovitch et al., 2003).

Without RbpA or CarD present, the mycobacteria holoenyzme is comparatively weaker than the *Eco* enzyme, partially a result of unstable intermediates and RPo (this work, Davis et al., 2014; Rammohan et al. 2015; 2016). Observations and experimental evidence that *Bacillus subtilis* and *Taq* RNAP also form less stable RPo (Whipple and Sonenshein, 1992; Schroeder and deHaseth, 2005; Miropolskaya et al., 2012) have led others to propose that these LSIs may play a role in RPo formation (Artsimovitch et al., 2003). The structure of Msm allows us to investigate whether this instability could be a result of structural differences between *Msm, Eco,* and *Taq* RNAPs. We therefore aligned the *Eco* RNAP structure (Bae et al., 2013 4LJZ) and the *Taq* holoenzyme (Bae et al., 2015; 4XLR) to our current *Msm* initiation complex (Chapter 3) and compared the sites of the lineage specific inserts (**Figure 6.1**).

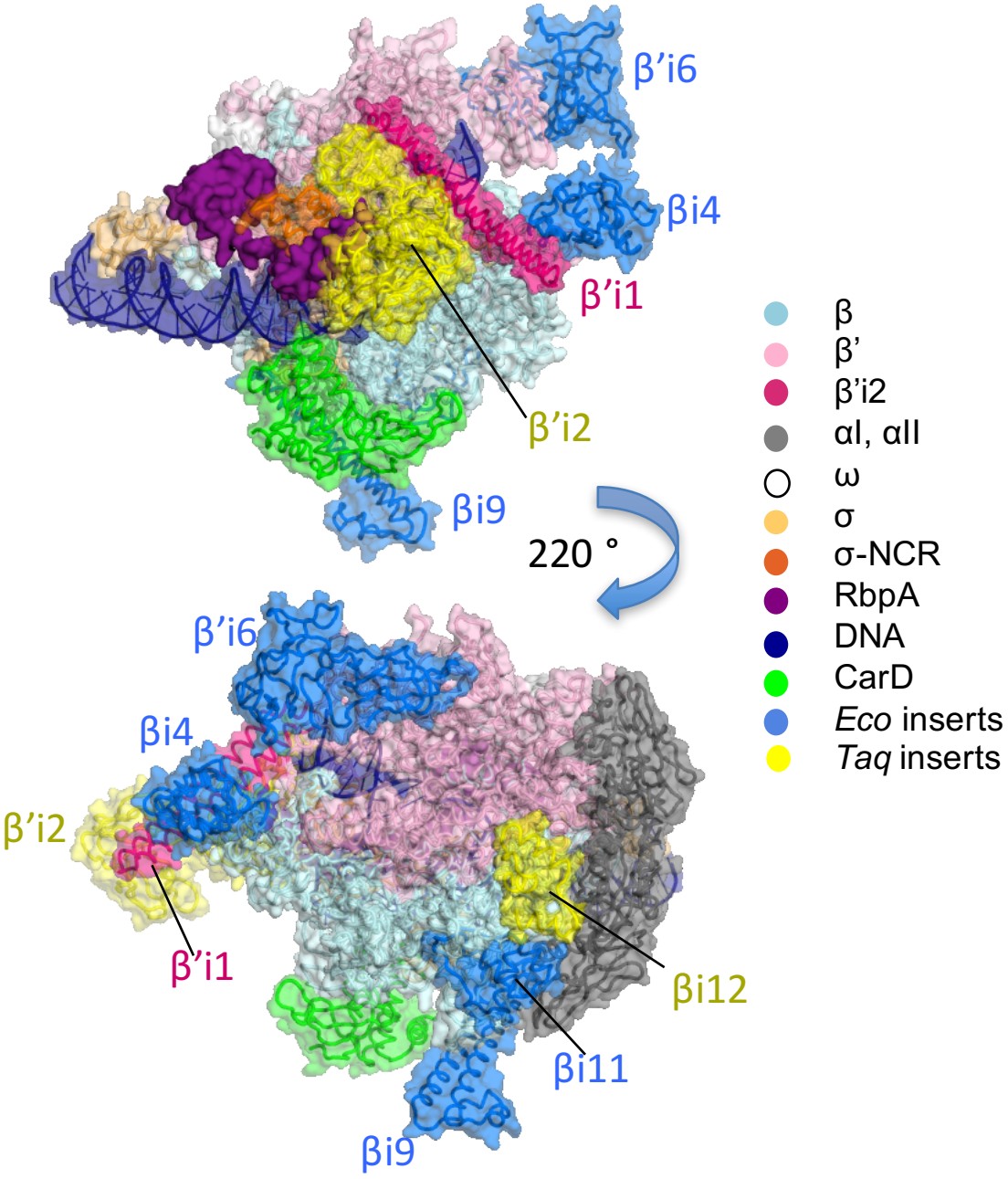


Figure 6.1 Lineage specific inserts from *Taq* (yellow) and *Eco* (blue) modeled on the *Msm* initiation complex with RbpA and CarD. Proteins and DNA colored as indicated.

Eco has four large inserts, βi4, βi9, βi11 and β'i6 (Lane and Darst, 2010a; 2010b; Opalka et al., 2010) that are not present in mycobacteria or *thermus,* and two of which (βi4 and β'i6) have been shown to contribute to RPo stability *in vitro* (Artsimovitch et al., 2003). βi4 is inserted within the β-lobe 2, close to downstream DNA, exiting the main channel. Bacteria harboring a βi4 deletion show mild deficiency in RPo formation, sensitivity to temperature, and slow growth (Artsimovitch et al., 2003). β'i6 is inserted into the trigger loop, and deletions led to cell death. In vitro assays have revealed that RNAP with a deletion of β'i6 show reduced RPo stability and pausing (Artsimovitch et al., 2003).

Less is known about the function of the other two inserts: βi9 is attached to the flap domain by flexible linkers, is highly flexible in position, and could theoretically make contacts with DNA upstream of the -10. Deletion mutants lacking the βi9 insert behave similarly to wild-type RNAP in RPo stability and pausing assays; however, this deletion results in bacteria that cannot support growth in minimal media (Artsimovitch et al., 2003). Located at the back of the *Eco* enzyme near the α-NTD dimer is βi11, which was just recently reported as a LSI; hence there are no published studies of its function.

Taq contains two LSIs: βi12, inserted near the a-NTD dimer, close to the location of βi11 in *Eco*; and β'i2, inserted on the β' clamp domain close to σ2 (**Figure 6.1**). Interestingly, β'i2 is inserted into the same position on the β' clamp as β'i1, the only significant insert in mycobacteria. β'i1, specific to actinobacteria and a few other clades, is revealed by our structure to be composed of two long parallel helices. The significance of the location of β'i2 and β'i1 is unknown; however, we are testing its importance in mycobacteria.

The comparison of structures and locations of the LSIs between *Eco*, *Taq*, and *Msm* does not offer an obvious explanation of the differences in RNAP activity between these organisms. We note that some of these inserts are highly mobile, and their positions could vary during the steps of initiation, promoter escape, pausing, elongation and termination. The functional importance of the LSIs and their roles in transcription remains relatively unexplored, especially within organisms other than *Eco*. It is worth noting that the two inserts (β'-i6 and β-i4) in *Eco* that are required for promoter stability are absent in mycobacteria and also extend the pincers of the clamp near the downstream DNA. It is possible these inserts affect the position of the clamp in RPo, stabilizing their interaction with the downstream DNA.

The *Msm* holoenzyme structure revealed that in the context of an initiation complex, region 1 of σ ($\sigma_{1.1}$, described in Chapter 3) is located near β-lobe 2 (close to where Eco βi4 would be) and is within 4 Å of *Msm* LSI β'i1 (**Figure 6.2)**. It is also close to where *Taq* β'i2 would be (**Figure 6.2)**. With their proximity to $\sigma_{1.1}$, it is possible that these LSI's may play a role in RPo formation by aiding in the displacement of $\sigma_{1.1}$ during the placement of the downstream DNA.

The NCR of σ^A from *Taq*, σ^{70} from *Eco,* and σ^A from *Msm* differ significantly in sequence, size and structure (**Figure 6.3**). While the NCR of σ^A is the smallest of the three, it likely that σ is almost always bound to RbpA in the cell, essentially extending its NCR. Notably, neither the NCR from *Taq* σ^A nor *Eco* σ^{70} overlap with any region of RbpA or make contacts with DNA, so it is unlikely that they serve a function similar to RbpA.

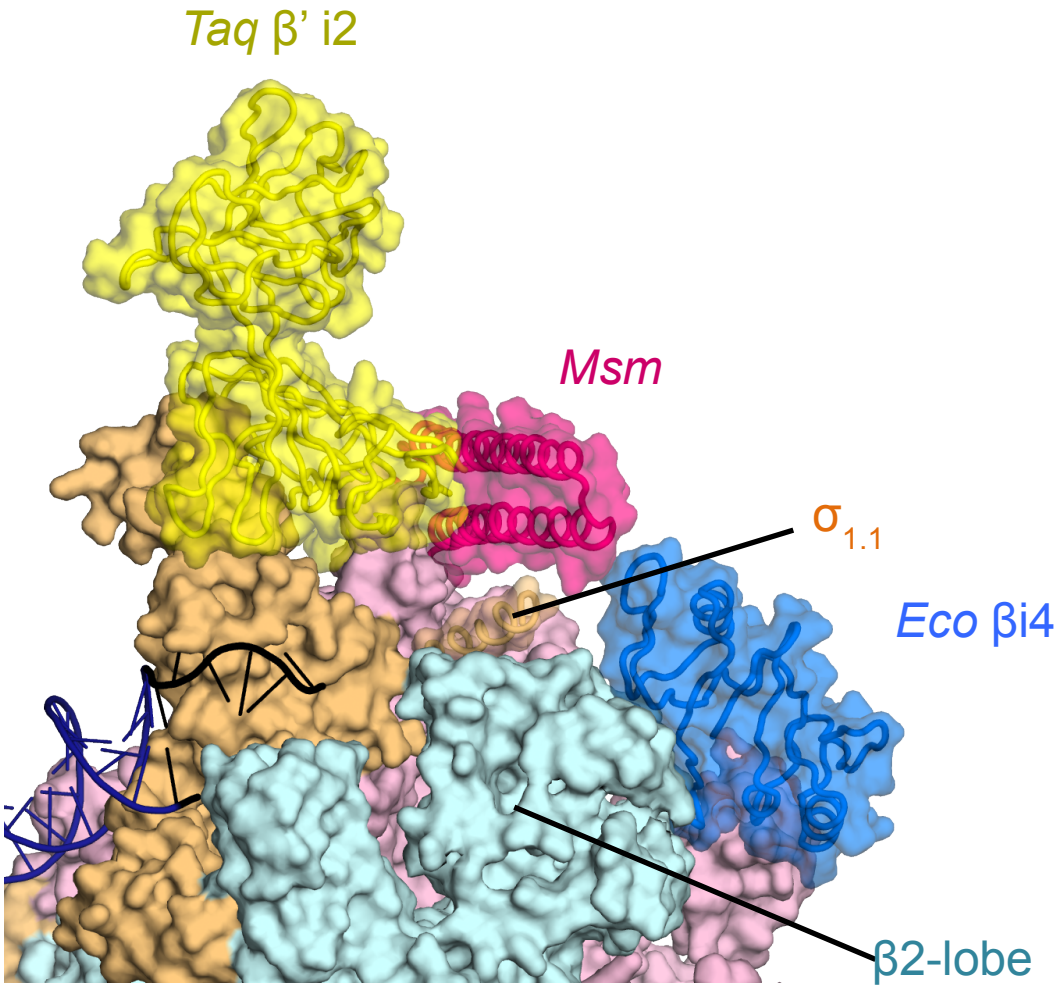


Figure 6.2: $\sigma_{1.1}$ is in close proximity to LSIs in *Taq*, *Eco*, and *Msm*. βi4 and β'i2 aligned on the *Msm* holoenzyme structure. $\sigma_{1.1}$ is located between β' and the protrusion, comes within 4 Å of *Msm* LSI β'i1, and is located near *Taq* β'i2 and *Eco* βi4.

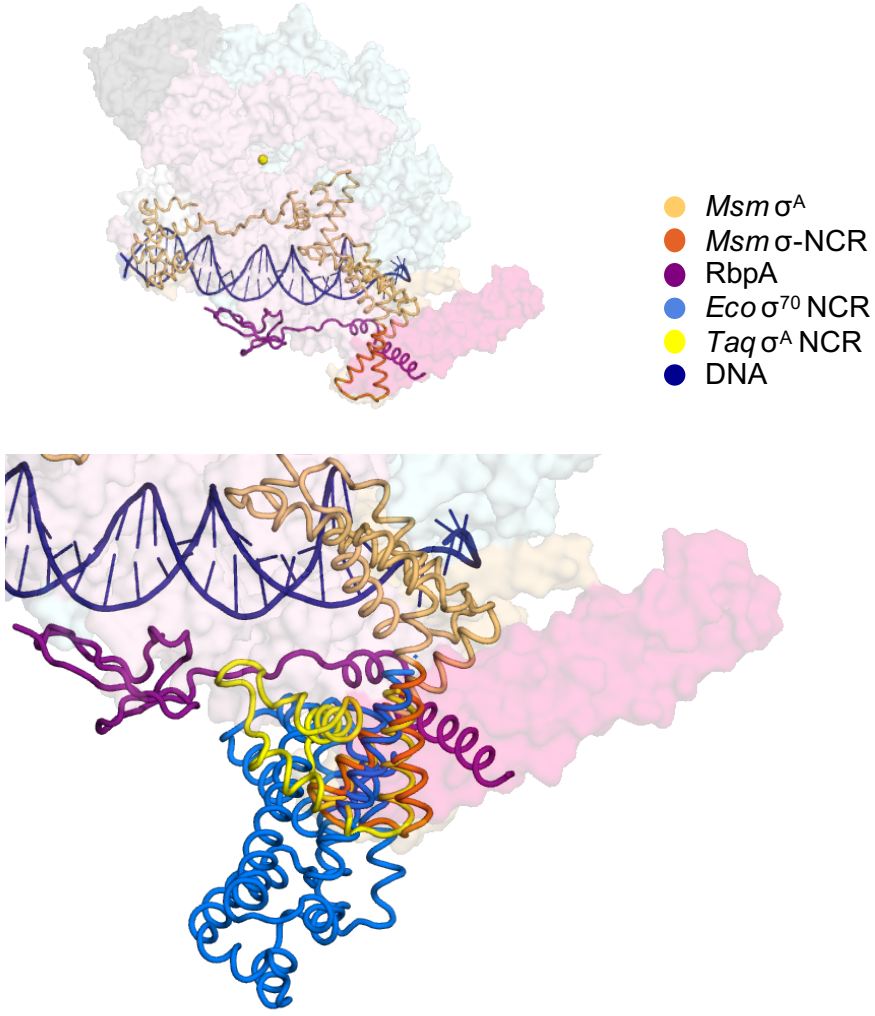


Figure 6.3: NCR of σ^A from *Taq*, σ^{70} from *Eco* modeled on *Msm* holo-RbpA. Top: *Msm* initiation complex bound to RbpA. View without *Taq* and *Eco* NCRs for orientation. Bottom: NCR of σ^A from *Taq* and σ^{70} from *Eco* modeled on *Msm* holo-RbpA and zoomed in to orientation shown above. Proteins and DNA colored as indicated.

Chapter 7:

Conclusions and Future Directions

The body of work presented in this thesis comprises a structural, biochemical, and kinetic characterization of RbpA, an essential transcription factor in the pathogen *Mtb* (Forti et al., 2011). Structures of the *Mtb* RbpA-SID/σ^{A}_{2} complex and the *Msm* full initiation complex bound to RbpA revealed that the BL (specifically conserved residue RbpA-R79) contacts promoter DNA upstream of the transcription bubble. Further biochemical analysis revealed this interaction to be the basis for RbpA's mechanism for transcription activation. A kinetic assay measuring RPo formation revealed that the contacts between the RbpA-BL and the DNA upstream of the bubble drives the forward rate of formation of an RNAP-promoter intermediate, which, in turn, increases populations of RPo.

A synergistic effect between RbpA and the transcription factor CarD (also essential in *Mtb*) is striking in both abortive initiation half-life assays and in kinetic assays that measure RPo formation. This effect indicates that they work together to activate transcription in the cell. However, additional studies should confirm that RbpA and CarD co-localize on the same RNAP *in vivo.*

Future studies will focus on the function of the RbpA CD and NTD, which are not required for full activation by RbpA *in vitro,* but are necessary for normal growth *in vivo*. That the CD represses transcription activation is especially interesting and possibly reflects a function of RbpA other than driving RPo formation. The *Msm* holo-RbpA structure reveals the CD contacts zipper and ZBD, which may provide clues to its function. The zipper region has been shown to play a role in σ-dependent pausing

(Yuzenkova et al., 2011): thus future studies will test the effect of the RbpA truncations on σ-dependent pausing. The zipper has also been shown to play a role in binding promoter DNA within the spacer region called the "Z-element" (Yuzenkova et al., 2011). While the CD does bind the zipper in this region, it may have an allosteric effect on the zipper contacting promoters with specific z-element sequences. The function of the NTD is also unknown, as this region is quite disordered and flexible, and the effect of the $RbpA^{\Delta NTD}$ mutant seems to be slightly variable between different assays and promoter types.

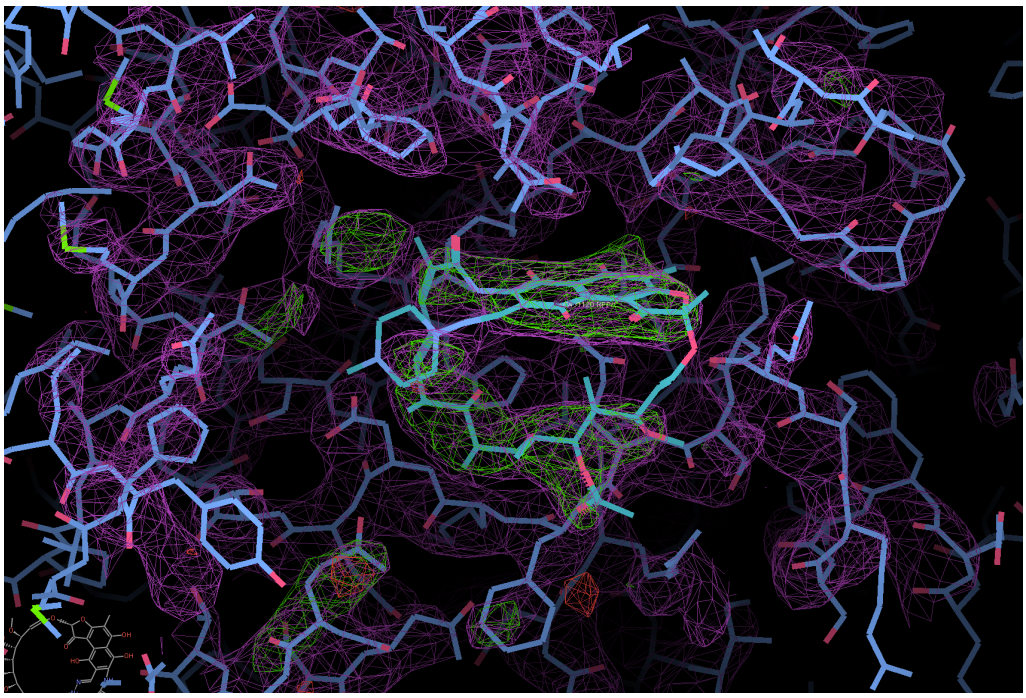
To date, transcription activation by RbpA has been measured only on a handful of promoters, and it is possible that the function of the CD or NTD depends on the architecture and kinetics of specific promoters. Therefore, in collaboration with Mike Glickman's lab, we are performing RNA-seq on the *Msm* cell lines with WT RbpA and with those harboring the RbpA truncation mutations. Comparison of the data may reveal differential gene regulation in the context of the truncations, which would provide insight into the whether the CD and NTD are necessary on certain promoters.

The crystal structure of *Msm* RNAP presented in this thesis has provided a platform for structure-based analysis of current and novel RNAP inhibitors that inhibit the growth of *Mtb*. The Darst lab is currently using the *Msm*-holo-RbpA-T10 crystal form (detailed in Chapter 3) to study structures of *Msm* RNAP bound to rifampicin and sorangicin, an RNAP inhibitor that binds to the same site and works by a similar mechanism as rifampicin (Campbell et al., 2005). Nathaniel Braffman in the Darst lab has obtained structures to 3.0 Å containing density for each antibiotic (**Figure 7.1**). Future studies will use this crystal form to study novel RNAP inhibitors, including derivatives of

rifampicin in order to understand their mechanism of inhibition of mycobacteria RNAP with atomic detail.

Both *Msm* RNAP crystal forms presented in this thesis are unable to accommodate the binding of CarD. Thus future studies will pursue additional crystal forms or use cryo-EM to generate a mycobacteria structure with both RbpA and CarD bound. Such a structure would represent a "complete" mycobacterial initiation complex.

A



B

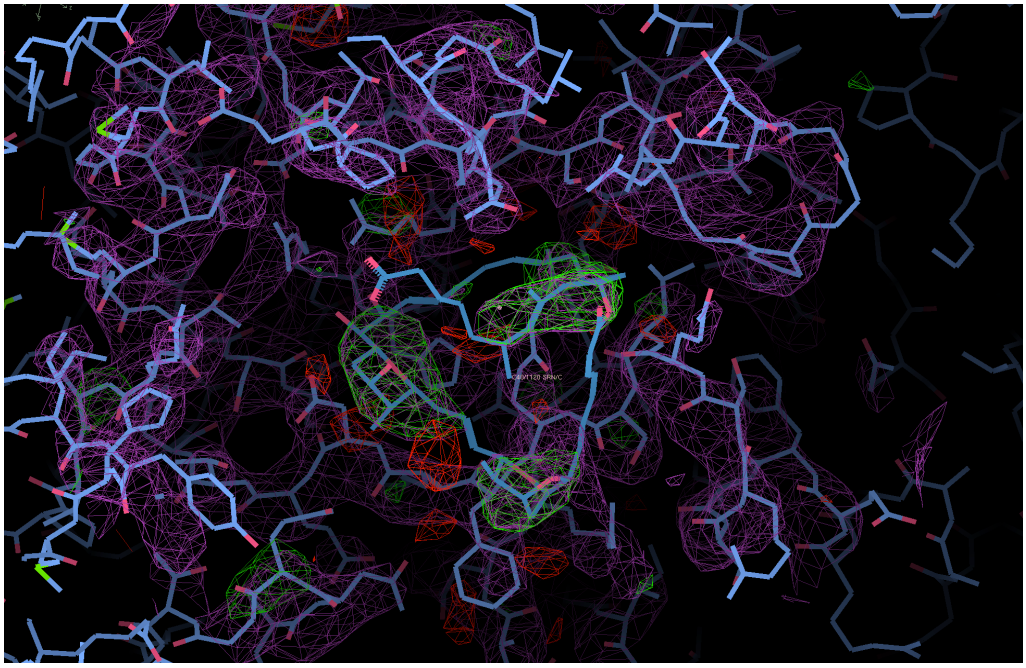


Figure 7.1: Structures of *Msm* holo-RbpA-T10 bound to antibiotics determined to 3Å. 2FoFc is contoured to a sigma of 1; FoFc map is contoured to a sigma of 3. Electron density maps show FoFc positive (green) density for (A) Rifampicin (B) Sorangicin

Chapter 8:

Materials and Methods

Protein expression and purification:

For crystallization of the σ^{A}_{2}/RbpA complex, σ^{A}_{2} (codons 222-364) and *rbpA* (codons 1-111) from *Mtb* were chemically synthesized as a single DNA fragment with *Eco* codon usage and independent T7 promoter and translation initiation signals for *rbpA* (Genscript), and then cloned as a *Bam*HI-*Hin*dIII fragment into a His_6 pET SUMO expression vector and transformed into *Eco* BL21(DE3) cells (Novagen). Transformed cells were grown at 37 °C in the presence of 50 μg/ml kanamycin to an OD_{600} of 0.6, at which point the temperature was lowered to 30 °C, and protein expression was induced with 500 μM IPTG for 3 h. Cells were harvested by centrifugation and resuspended in Lysis Buffer [20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole, 5% glycerol (v/v), 0.5 mM beta-mercaptoethanol] supplemented with 1 mM PMSF and protease inhibitor cocktail. Cells were lysed by French press (Avestin), and lysate was cleared by centrifugation. Clarified lysate was loaded on a Hi-Trap IMAC Ni^{2+}-chelating column (GE Healthcare) and eluted with Lysis Buffer containing 250 mM imidazole. The elution was directly loaded on a size exclusion column (SuperDex-200 16/16, GE healthcare) equilibrated with 20 mM Tris-HCl, pH 8, 0.5 M NaCl, 5% glycerol (v/v), 1 mM DTT, 1 mM EDTA. The sample was concentrated to 15 mg/ml by centrifugal filtration and stored at -80 °C prior to crystallization.

For crystallization of *Msm* holo-RbpA, *Msm* pET-SUMO σ^{A} (received from Christina Stallings' lab) and pet21C RbpA (cloned from *Msm* genomic DNA) were co-

transformed into BL21 *Eco* RNAP. Protein was expressed and purified by nickel column using the same buffers and techniques as the *Mtb* RbpA/σ_2 co-expressed complex. Following the nickel column fractionation of the complex, imidazole was removed by o/n dialysis back into Lysis Buffer, and the his-SUMO tag was cleaved σ^A by ULP1 protease post cleavage of the his-SUMO tag from σ^A. The complex was loaded onto a second nickel column and was retrieved from the flow-through and low-imidazole washes. The complex was directly loaded on a size exclusion column (SuperDex-200 16/16, GE healthcare) equilibrated with 20 mM Tris-HCl, pH 8, 0.5 M NaCl, 5% glycerol (v/v), 1 mM DTT, 1 mM EDTA. The sample was concentrated to 15 mg/ml by centrifugal filtration and stored at -80 °C.

Msm RNAP was purified endogenously from an *Msm* mc1255 strain expressing a native chromosomal copy of *rpoC* with a C-terminal 10-his tag (Generated by Mike Glickman's lab). *Msm* cells were grown at the Bioexpression and Fermentation Facility at the University of Georgia. Cells were lysed by French press in lysate buffer [50 mM Tris-HCl (pH 8), 1 mM EDTA, 5% Glycerol, 5 mM DTT, 1mM protease inhibitor cocktail, 1mM PMSF] and core RNAP was precipitated from cleared lysate by Polyethyleneimine precipation and eluted in PEI elution buffer [10 mM Tris-HCl (pH 8), 5% Glycerol, 0.1 mM EDTA, 5 mM DTT, 1M NaCl]. Protein was precipitated overnight with ammonium sulfate and resuspended in Nickel Buffer A [20 mM Tris, 5% glycerol, 1 M NaCl, 1 mM BME]. Protein was loaded on nickel column and eluted in Nickel Elution Buffer [20 mM Tris, 5% glycerol, 500 mM NaCl, 250 mM imidazole]. Protein was diluted in no-salt buffer [10 mM Tris-HCl (pH 8), 5% Glycerol, 0.1 mM EDTA, 5 mM DTT] to a final salt concentration of 100 mM NaCl and purified by a Biorex ion exchange column A. *Msm*

Holoenzyme RNAP containing *Msm* σ^A and RbpA was assembled by adding 5 molar excess of *Msm* σ^A/ RbpA protein and then purified as a complex by size exclusion chromatography (SD 200, GE healthcare) in gel filtration buffer [20 mM Tris, 5% glycerol, 500 mM NaCl, 250 mM imidazole]. Protein was dialyzed [100 mM KGlu, 20 mM Tris HCl pH8, 10 mM MgCl2, 1 mM DTT], concentrated by centrifugal filtration to ~15 mg/ml and stored at -80 °C.

Mtb core was purified using methods described by Davis et al., 2015. Briefly, *Mbo* core RNAP subunits were co-expressed from pAC22 (gifted from Robert Landick's lab; Czyz et al., 2014) in *E. coli* pRARE3 (Novagen) at room temperature for ~16 hr after induction with IPTG. Cell pellets were lysed and cleared lysate was treated with PEI. Elution was purified by Ni^{+2} chromatography followed by size exclusion chromatography. To obtain *Mbo*-holo-RbpA for crystallography, prior to gel purification, a 5-fold excess of *Mbo* σ^A/RbpA was added. *Mbo* σ^A/RbpA was prepared by co-transformation of *Mbo* pRM629 σ^A (gifted from Robert Landick's Lab; Czyz et al., 2014) with *Mtb* pet20b-RbpA into *E. coli* pRARE3 (Novagen). A complex was co-expressed and purified by the same methods used to purify *Msm* σ^A/RbpA.

For fluorescence anisotropy, in vitro transcription assays, and the Heyduk assay, *Mtb* RbpA-R79A was purified from at pET28 based vector using the same buffers and methods for the $\sigma^A{}_2$/RbpA complex. His_6-SUMO RbpA (amino acid residues 1-111) and truncated derivatives (amino acid residue 1-71 and 72-111) were cloned by Mark Paget, and purified using the same expression protocol and purification methods used to purify the to *Msm* σ^A/RbpA complex*:* Ni-affinity, reverse nickel followed by cleavage of SUMO tag, and size exclusion chromatography (HiLoad 16/60 Superdex 200, GE

Healthcare were cloned. Core *Mbo* RNAP, *Mbo* σ^A, and *Mtb* CarD were over-expressed and purified using previously described method (Davis et al., 2015).

Crystallization of RbpA/σ^A_2 complex

Crystals of the RbpA/σ^A_2 complex were grown at 22 °C via sitting drop vapor diffusion against a reservoir solution of 0.1 M Tris, pH 8.5, and 0.5 M ammonium sulfate, at a protein concentration of 15 mg/ml. Drops were set up with a 1:1 ratio of His_6-SUMO-σ^A_2-RbpA:crystallant. Crystals measuring approximately 50 μm took 3 days to grow and were cryoprotected in reservoir solution plus 30% (v/v) ethylene glycol.

Data collection and refinement of σ^A_2-RbpA complex:

X-ray diffraction data were collected at the Brookhaven National Synchroton Light Source X29 beamline and integrated and scaled using HKL2000 Electron density maps were generated using molecular replacement with an *Mtb* σ^A_2 homology model (ExPASy SWISS MODEL) based on a previously solved structure of the corresponding domain from *Taq* σ^A [1KU2; (Campbell et al., 2002)]. The model was built using reiterative cycles of manual building with COOT (Emsley and Cowtan, 2004)and refinement with Phenix (Adams et al., 2010) (Table S2). The final model included residues 242-363 of σ^A and 77-108 of RbpA. The RCD of RbpA, which lacked density, was presumed disordered. The PDBePISA server (http://www.ebi.ac.uk/pdbe/pisa/) was used to calculate intermolecular buried surface areas.

Fluorescence anisotropy

Fluorescence measurements were performed using an Infinite M1000pro plate reader (Tecan) in a 384 well plate with a final reaction volume of 25 uL. Holoenzyme was formed at 37°C using a 1:1.5 ratio of core RNAP:σ^A. All proteins were dialyzed and then serially diluted in Anisotropy Buffer consisting of 100 mM potassium-glutamate, 10 mM HEPES, pH 8, 10 mM $MgCl_2$, 0.1 mM DTT, 0.02% TWEEN 20, and 5% (v/v) glycerol. When used, RbpA was added to RNAP in 5-fold excess. Protein complexes were serially diluted in Anisotropy Buffer to obtain RNAP concentrations ranging from 5 nM to 3 μM. For assays with Cy3-labled single stranded DNA extending from the -12 to +1 (Tri-link, Table S3), DNA was diluted in Anisotropy Buffer and added to the protein mixture at a final concentration 200 nM. For assays performed with double stranded *vapB10* DNA extending from the -36 to +10, Cy3-labeled non-template strand DNA (Tri-link) was annealed to unlabeled template strand DNA (Oligos Etc.), diluted in Anisotropy Buffer, and added to the protein mixture at a final concentration of 10 nM. Measurements of RNAP binding to double-stranded DNA was noisy without the addition of CarD (a result of the unstable open complex and short half-life of *M. bovis* RNAP (Davis et al., 2015)), so these assays included CarD (in 3-fold excess over RNAP). Data were analyzed using Prism software. Assays with single-stranded DNA were performed in triplicate. For assays with duplex DNA, three experiments composed of side-by-side triplicate trials for each condition were performed.

Crystallization and *Mbo* holo-RbpA-T6 fork complex:

Crystals of the *Mbo* complex were grown at 22 °C via sitting drop vapor diffusion against a reservoir solution of 0.2 M ammonium sulfate, bis-tris pH 6.0 19% PEG 3350, at a protein concentration of 11 mg/ml. Drops were set up with a 1:1 ratio of protein:crystallant. Crystals measuring approximately 50 μm took 7 days to grow and were cryoprotected in reservoir solution plus 15% (v/v) ethylene glycol. Crystals were evaluated for diffraction at the APS NE-CAT Beamline in Chicago.

Crystallization and data collection of *Msm* holo-RbpA fork complex:

Crystals of the Msm holo-RbpA-T5 fork complexes were grown at 4 °C via sitting drop vapor diffusion against a reservoir solution of 0.25M $LiSO_4$, 0.1 M Bis-Tris pH 6.0, 18% PEG 3350, 5% Ethylene Glycol at a protein concentration of 11 mg/ml. Crystals were cryoprotected in 15% Ethylene Glycol. Crystals *Msm* holo-RbpA-T10 fork were grown at 22 °C via sitting drop vapor diffusion against a reservoir solution of 0.20M $LiSO_4$, 0.1 M Bis-Tris pH 6.0, 16% PEG 3350, 2.5% Ethylene Glycol at a protein concentration of 11 mg/ml. Crystals were cryo-protected (0.20M $LiSO_4$, 0.1 M Bis-Tris pH 6.0, 22% PEG 3350, 20% EG) and were gradually soaked into the cryo over three steps. Data collections of crystals were collected at the APS NE-CAT Beamline in Chicago. Data were indexed and scaled using HKL2000 (Krissinel and Henrick, 2007). Molecular replacement solutions were identified using phaser (Adams et al., 2010) and models were build using reiterative cycles of manual building with COOT (Emsley and Cowtan, 2004) and refinement with Phenix (Adams et al., 2010) The PDBePISA server

(http://www.ebi.ac.uk/pdbe/pisa/) was used to calculate intermolecular buried surface areas.

Heyduk assay:

Preparation of DNA: AP3 promoter DNA (-x to x) with a Cy3 label at the +2 position of the non-template strand was prepared using Cy3-amido-dT modified AP3 DNA (5'CATCTATGGATGACCGAACCTGGTCTTGACTCCATTGCCGGATTTGTATTA GACTGGCAGGGTTG/ICY3N/TG 3')

ordered from ITD. Cy3-labeled oligo (0.25 uM) with was mixed with template strand DNA (0.25 uM, Oligos etc.) containing a 20-bp complementary sequence at the 3' end (5'TTCTGAGTTCGGCATGGGGTCAGGTGGGACCCAAGCTTCCGCTTCGGGGCA ACCCTGCCAGTCTAATAC 3').

Taq DNA extention was performed on the partial duplex DNA, and the resulting product was visualized on a native gel stained with Gel Red to verify a single product, and then purified using PCR clean up.

VapB promoter DNA (-x to x) with a Cy3 label at the +2 position was prepared as above with the a VapB Cy3-labeled non-template strand (5'GGCTGAATCGCCGCCCGCCGCGGTGCCGCCCGGGCCGCACATTGTGATGT ATGATATGGTGTA/ICY3N/GA 3')

from IDT and unlabeled template strand from Oligo's etc. (5'CGTGTAACACTACATACTATACCACATACTTCGCCTGGTTGTAGATGGAGC TGCTCCTCGTCTGCCGTTCGGAGCTGTT 3')

Heyduk stopped-flow assay:

To monitor change in fluorescence, proteins (RNAP and transcription factors if present) were loaded in one syringe of a stopped flow instrument (KSF-300X, KinTek Corporation, Austin, Texas) and Cy3-labeled DNA was loaded into the other. Final buffer conditions for all reactions were: 50 mM KGlu, 20 mM NaCl, 10 mM MgCl2, 1 mM DTT. For AP3, final concentration of DNA was 1 nM and for VapB, DNA was 3 nM. RNAP concentration ranged from 3 nM-400 nM for AP3 and 10 nM-500 nM for VapB. When present, RbpA (and RbpA derivatives) and CarD were kept at saturating concentrations (5 uM and 10 uM, respectively). Reactions were performed at 37 °C. Reactions were excited at 515 nM, and change in Cy3 emission was measured in real time with a 586/20 single bandpass filter.

Data-fitting for Heyduk assay

Data for the Heyduk assay on AP3 and VapB were fit using KinTek Global Explorer ©. Data best fit a 3-step linear model for RPo formation (Section 5.2; **Appendix 2**), with the observables a*[P]) + (b*[RP1]) + (c*[RP2]) + (d*[RPo]) + bkg. Concentration series collected on separate days were fitted separately, always yielding similar constants. The constants derived in this way were then averaged, giving rise to the average values and standard error.

For AP3, data was refined by constraining values for k_{-3} and k_3: for each reaction k_{-3} was adjusted and then fixed so that the theoretical half-life matched experimental data; k_3 was fixed so that it was the same for every reaction. AP3 data was then validated using KinTek Fitspace. For the fitted kinetic rate constants are based on the standard error of

the average of multiple experiments for each rate constant, with the exception of the *Eco* holo sample, which was performed only once, and thus errors are based on calculated errors from the KinTek explorer. Equilibrium constants were calculated with averaged rate constants (for example K1=averaged k_1/averaged k_{-1}) and error was based on a propagation of error. For *Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+CarD+RbpA, three experiments were averaged; for *Mbo* holo+CarD, 2 experiments were averaged.

For VapB, two experiments, on separate days, were performed for all samples (*Mbo* holo, *Mbo* holo+RbpA, *Mbo* holo+RbpA$^{SID\text{-}BL}$, *Mbo* holo+RbpA$^{\Delta NTD}$, and *Mbo* holo+RbpA^{R79A}) and rate constants and equilibrium constants were calculated as they were for AP3.

Activation energy for each step was calculated using the Global KinTek Explorer based on the fitted rate constant values using the equation:

$$\Delta G_i^{\dagger} = RT \ln\left(\frac{k_B T / h}{k_i}\right)$$

Where $\Delta G^{\ddagger}_i$ = activation energy, $h = 6.62e^{-34}\ J{\bullet}s$, $k_B = 1.38e^{-34}\ J / K$, and k_i = rate constant variable.

Transcription assays

AP3 promoter was prepared as outlined by Davis et al., 2014. To prepare VapB promoter used for transcription assays, p*vapB10L* promoter template was amplified by PCR using the oligonucleotides vapB10_F (5'-GCGCTGAAGAGGGCGTTGCAC) and vapB10_R (5'-TTCAGCAGGAGGCGGATCAG). PCR products were run on a 1.5% agarose gel and purified by standard gel purification methods.

Abortive initiation assays and run-off assays were performed at 37 °C. Assays performed on $AP3^{\Delta-35}$ were in KGlu buffer (10 mM Tris-HCl, pH 8.0, 100 mM potassium-glutamate, 10 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, and 50 μg/mL BSA). All other assays were performed in KCl buffer (10 mM Tris-HCl, pH 8.0, 50 mM potassium chloride, 10 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, and 50 μg-/mL BSA). *Mbo* core (50 nM) was incubated with σ^A (250 nM) for 5 min to form holoenzyme, and when used, CarD and or RbpA/RbpA derivatives (all 2 μM) were added and incubated for an additional 5 min. DNA (10 nM) was then added and incubated with the protein mixture for 15 min to form RPo. Abortive transcription for abortive initiation assays were initiated on VapB with an ApU dinucleotide primer (250 μM), [α-^{32}P]GTP (1.25 μCi), and unlabeled GTP (50 μM). Abortive transcription for abortive initiation and half-life assays were initiated on pAP3 with GpU dinucleotide primer (250 μM), [α-^{32}P]UTP (1.25 μCi), and unlabeled UTP (50 μM). After 10 min, reactions were quenched with 2x stop buffer (8M urea, 0.5x TBE, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on a 23% urea-polyacrylamide gel. Abortive products were visualized by phosphorimagery and quantified using Image J. For run off assays on AP3 and vapB, transcription was initiated with an NTP mix containing unlabeled GTP, ATP and CTP (250 μM), unlabeled UTP (50 μM) and [α-^{32}P]UTP (1.25 μCi). After 10 min, reactions were quenched with 2x stop buffer (8M urea, 0.5x TBE, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on a 8% urea-polyacrylamide gel. Run-off products were visualized by phosphorimagery and quantified using Image J.

Abortive initiation half-life assays on AP3 were performed in buffer (10 mM Tris-HCl, pH 8.0, 10 mM KGlu, 10 mM $MgCl_2$, 0.1 mM DTT, and 50 μg/ml BSA), at 37 °C.

Holoenzyme was formed by mixing core (200 nM) and σ^A (1 μM) and incubating for 5 min. Transcription factors (2 μM) were then added and the reaction was incubated for an addition 5 min. DNA (10 nM) was added, and RPo was allowed to form for 15 min. FC competitor was incubated with the reaction for various times (0, 1, 5, 16, 30, 60, 120, 180 min) before initiating with GpU dinucleotide primer (250 μM), [α-^{32}P]UTP (1.25 μCi), and unlabeled UTP (50 μM). After 10 min, reactions were quenched with 2x stop buffer (8M urea, 0.5x TBE, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on a 23% urea-polyacrylamide gel.

Appendix 1:

Sequences of promoters used to study mycobacterial transcription

Cyan: Initiating 3 nucleotides
Pink highlighted: 35 element
Yellow highlighted: -10 element
Red highlighted: discriminator
Bold: +1 base

Non-template strand sequence:

VapB

GCGCTGAAGAGGGCGTTGCACGCATAACGTCGGCGGCATGCCCGGTCGGCTGAATCGCC
GCCCGCCGCGGTGCCGCCCGGGCCGCACATTGTGATG**TATGAT**ATGGTGTATGAAGCGG
ACCAACATCTACCTCGACGAGGAGCAGACGGCAAGCCTCGACAAGTTGGCCGCGCAAGA
AGGTGTTTCGCGCGCCGAGCTGATCCGCCTCCTGCTGAA

Mtb AP3

GGATCCCCCGAAAAGTGCCACCTGACGGAATTCATCTATGGATGACCGAACCTGGTCTT
GACTCCATTGCCGGATTTGTATTAGACTGGCAGGGTTGCCCCGAAGCGGAAGCTTGGGT
CCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGGATCC

Mtb AP3Δ-35

GGATCCCCCGAAAAGTGCCACCTGACGGAATTCATCTATGGATGACCGAACCTGGTCAC
TCAGCCATTGCCGGATTTGTATTAGACTGGCAGGGTTGCCCCGAAGCGGAAGCTTGGGT
CCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGGATCC

AC50

CCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAA
AATAGGCGTATCACGAGGCCCTTTCGTCTTCAA**GAATTC**AAATATTTGTTGTTAACTCT
TGACAAAAGTGTTAAAAGCGGCTAGTATTTAAAG**G**GATGGATGACATCTC**AAGCTT**GGG
TCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCC

Appendix 2:

Tables of kinetic values for various models fitted to the Heyduk assay on AP3

sample	M. bovis holo				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(1.1 ± 0.4) x 10^6	(7.8 ± 1.2) x 10^6	(9.5 ± 1.3) x 10^6	(9.5 ± 1.5) x 10^6	(9.9 ± 1.7) x 10^6
k_{-1} (s^{-1})	(6.0 ± 2.2) x 10^{-3}	2.0 ± 0.2	2.0 ± 0.2	1.6 ± 0.2	1.6 ± 0.3
K_1 (M^{-1})	(1.8 ± 0.9) x 10^8	(3.9 ± 0.7) x 10^6	(4.8 ± 0.8) x 10^6	(5.9 ± 1.2) x 10^6	(6.2 ± 1.6) x 10^6
k_2 (s^{-1})		0.37 ± 0.02	0.40 ± 0.02	0.12 ± 0.04	0.3 ± 0.07
k_{-2} (s^{-1})		(9.5 ± 2.3) x 10^{-3}	0.058 ± 0.007	0.21 ± 0.04	0.097 ± 0.021
K_2		39 ± 10	6.9 ± 0.9	0.57 ± 0.22	3.1 ± 1.0
k_3 (s^{-1})			0.057 ± 0.011	0.18 ± 0.03	0.21 ± 0.16
k_{-3} (s^{-1})			0.015 ± 0.004	(6.8 ± 3.7) x 10^{-3}	2.2 ± 11.5
K_3			3.8 ± 1.3	26 ± 15	0.095 ± 0.504
k_4 (s^{-1})					3.8 ± 25.5
k_{-4} (s^{-1})					0.048 ± 0.105
K_4					0.025 ± 0.215
χ^2/DOF	1.62744	1.15668	1.06622	1.06659	1.06737
k_d[a]	6.0 x 10^{-3}	9.4 x 10^{-3}	6.6 x 10^{-3}	6.1 x 10^{-3}	6.0 x 10^{-3}
$t_{1/2}$ (min)	1.9	1.2	1.7	1.9	1.9
a	0.48	0.55	0.30	0.48	0.65
b	1.3	0.81	0.51	0.67	0.83
c		1.4	1.2	1.8	1.8
d			1.2	1.4	0.026
e					1.5
bkg	1.3	0.16	0.41	.23	0.064

sample	M. bovis holo				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(1.1 ± 0.4) x 10^6	(7.8 ± 1.2) x 10^6	(9.5 ± 1.3) x 10^6	(9.5 ± 1.5) x 10^6	(9.9 ± 1.7) x 10^6
k_{-1} (s^{-1})	(6.0 ± 2.2) x 10^{-3}	2.0 ± 0.2	2.0 ± 0.2	1.6 ± 0.2	1.6 ± 0.3
K_1 (M^{-1})	(1.8 ± 0.9) x 10^8	(3.9 ± 0.7) x 10^6	(4.8 ± 0.8) x 10^6	(5.9 ± 1.2) x 10^6	(6.2 ± 1.6) x 10^6
k_2 (s^{-1})		0.37 ± 0.02	0.40 ± 0.02	0.12 ± 0.04	0.3 ± 0.07
k_{-2} (s^{-1})		(9.5 ± 2.3) x 10^{-3}	0.058 ± 0.007	0.21 ± 0.04	0.097 ± 0.021
K_2		39 ± 10	6.9 ± 0.9	0.57 ± 0.22	3.1 ± 1.0
k_3 (s^{-1})			0.057 ± 0.011	0.18 ± 0.03	0.21 ± 0.16
k_{-3} (s^{-1})			0.015 ± 0.004	(6.8 ± 3.7) x 10^{-3}	2.2 ± 11.5
K_3			3.8 ± 1.3	26 ± 15	0.095 ± 0.504
k_4 (s^{-1})					3.8 ± 25.5
k_{-4} (s^{-1})					0.048 ± 0.105
K_4					0.025 ± 0.215
χ^2/DOF	1.62744	1.15668	1.06622	1.06659	1.06737
k_d [a]	6.0 x 10^{-3}	9.4 x 10^{-3}	6.6 x 10^{-3}	6.1 x 10^{-3}	6.0 x 10^{-3}
$t_{1/2}$ (min)	1.9	1.2	1.7	1.9	1.9
a	0.48	0.55	0.30	0.48	0.65
b	1.3	0.81	0.51	0.67	0.83
c		1.4	1.2	1.8	1.8
d			1.2	1.4	0.026
e					1.5
bkg	1.3	0.16	0.41	.23	0.064

sample	M. bovis holo + RbpA				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(6.4 ± 0.3) x 10^6	(1.2 ± 0.1) x 10^7	(1.5 ± 0.1) x 10^7	(1.5 ± 0.1) x 10^7	(1.6 ± 0.1) x 10^7
k_{-1} (s^{-1})	(6.5 ± 1.1) x 10^{-3}	0.80 ± 0.07	1.1 ± 0.1	1.1 ± 0.1	0.88 ± 0.09
K_1 (M^{-1})	(9.9 ± 1.7) x 10^8	(1.5 ± 0.2) x 10^7	(1.4 ± 0.2) x 10^7	(1.4 ± 0.2) x 10^7	(1.8 ± 0.2) x 10^7
k_2 (s^{-1})		1.1 ± 0.1	1.4 ± 0.1	0.53 ± 0.05	0.81 ± 0.08
k_{-2} (s^{-1})		0.017 ± 0.002	0.090 ± 0.020	(7.1 ± 2.6) x 10^{-3}	0.062 ± 0.030
K_2		65 ± 10	16 ± 4	75 ± 28	13 ± 6
k_3 (s^{-1})			0.046 ± 0.015	0.86 ± 0.05	0.59 ± 0.75
k_{-3} (s^{-1})			0.011 ± 0.004	0.14 ± 0.02	1.6 ± 0.7
K_3			4.2 ± 2.0	6.1 ± 0.9	0.37 ± 0.50
k_4 (s^{-1})					0.091 ± 0.193
k_{-4} (s^{-1})					(9.9 ± 14) x 10^{-3}
K_4					4.1 ± 10.2
χ^2/DOF	1.47786	1.24241	1.11268	1.11269	1.10116
k_d[a]	6.4 x 10^{-3}	0.016	6.4 x 10^{-3}	4.6 x 10^{-3}	3.6 x 10^{-3}
$t_{1/2}$ (min)	1.8	0.71	1.8	2.5	3.2
a	0.099	0.038	0.16	0.43	0.65
b	1.2	0.70	0.68	0.94	0.83
c		1.1	1.3	1.5	1.8
d			1.3	1.5	0.026
e					1.5
bkg	0.81	0.86	0.74	.47	0.064

sample	M. bovis holo + CarD				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(1.3 ± 0.1) x 10^7	(5.4 ± 0.4) x 10^7	(5.8 ± 0.5) x 10^7	(5.7 ± 0.5) x 10^7	(3.8 ± 0.3) x 10^7
k_{-1} (s^{-1})	(3.2 ± 1.4) x 10^{-3}	10 ± 1	6.2 ± 1.0	6.2 ± 1.0	1.4 ± 0.2
K_1 (M^{-1})	(4.1 ± 1.8) x 10^9	(5.4 ± 0.7) x 10^6	(9.4 ± 1.7) x 10^6	(9.2 ± 1.7) x 10^6	(2.7 ± 0.4) x 10^7
k_2 (s^{-1})		4.0 ± 0.3	3.1 ± 0.2	1.6 ± 0.2	1.1 ± 0.2
k_{-2} (s^{-1})		(2.7 ± 1.8) x 10^{-3}	0.067 ± 0.019	(1.8 ± 1.9) x 10^{-3}	0.049 ± 0.024
K_2		1500 ± 99	46 ± 14	890 ± 95	22 ± 12
k_3 (s^{-1})			0.074 ± 0.027	1.4 ± 0.1	0.72 ± 0.41
k_{-3} (s^{-1})			(2.5 ± 3.0) x 10^{-3}	0.16 ± 0.02	1.4 ± 0.4
K_3			30 ± 37	8.8 ± 1.3	0.51 ± 0.32
k_4 (s^{-1})					0.11 ± 0.06
k_{-4} (s^{-1})					(3.9 ± 5.3) x 10^{-3}
K_4					4.7 ± 3.9
χ^2/DOF	1.34903	1.22448	1.10877	1.10897	1.1021
k_d [a]	3.2 x 10^{-3}	2.6 x 10^{-3}	1.1 x 10^{-3}	1.4 x 10^{-3}	1.2 x 10^{-3}
$t_{1/2}$ (min)	3.6	4.3	10	8.3	9.6
a	0.50	0.56	0.28	0.48	0.80
b	1.5	0.75	0.45	0.66	1.1
c		1.5	1.2	1.4	2.6
d			1.2	1.4	5.3 x 10^{-5}
e					1.8
bkg	0.31	0.24	0.52	.32	2.9 x 10^{-3}

sample	M. bovis holo + RbpA + CarD				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(2.7 ± 0.1) x 10^7	(1.2 ± 0.2) x 10^8	(1.1 ± 0.2) x 10^8	(1.1 ± 0.2) x 10^8	(5.4 ± 0.3) x 10^7
k_{-1} (s^{-1})	(6.1 ± 1.5) x 10^{-3}	27 ± 5	20 ± 5	20 ± 5	2.0 ± 0.4
K_1 (M^{-1})	(4.4 ± 1.1) x 10^9	(4.4 ± 1.1) x 10^6	(5.5 ± 1.7) x 10^6	(5.5 ± 1.7) x 10^7	(2.7 ± 0.6) x 10^7
k_2 (s^{-1})		9.6 ± 0.4	9.7 ± 0.5	7.8 ± 0.5	3.9 ± 0.5
k_{-2} (s^{-1})		1.1 x 10^{-7}	0.031 ± 0.011	(2.1 ± 9.1) x 10^{-4}	0.11 ± 0.43
K_2		8.7 x 10^7	310 ± 11	(3.7 ± 16) x 10^4	35 ± 140
k_3 (s^{-1})			0.13 ± 0.03	1.9 ± 0.2	11 ± 34
k_{-3} (s^{-1})			(1.3 ± 8.2) x 10^{-3}	0.16 ± 0.03	9.0 ± 30
K_3			100 ± 63	12 ± 3	1.2 ± 5.6
					0.16 ± 0.42
					(2.9 ± 3800) x 10^6
					7.6 ± 40
χ^2/DOF	1.32367	1.09473	1.05035	1.05036	1.0458
k_d [a]	6.1 x 10^{-3}	1.1 x 10^{-7}	2.5 x 10^{-4}	3.4 x 10^{-7}	4.3 x 10^{-7}
$t_{1/2}$ (min)	1.9	1.1 x 10^5	47	3.3 x 10^4	2.7 x 10^4
a	0.59	0.63	0.28	0.28	0.91
b	1.7	0.63	0.28	0.28	0.95
c		1.6	1.2	1.3	4.2
d			1.3	1.1	5.7 x 10^{-6}
e					1.9
bkg	0.34	0.28	0.63	.63	1.1 x 10^{-6}

sample	Eco holo				
Model	1-step	2-step	3-step	3-step branched	4-step
k_1 ($M^{-1}s^{-1}$)	(9.9 ± 0.7) x 10^6	(1.2 ± 0.1) x 10^8	(1.2 ± 0.1) x 10^8	(1.2 ± 0.1) x 10^8	(8.5 ± 0.6) x 10^7
k_{-1} (s^{-1})	(2.7 ± 2.2) x 10^{-3}	4.6 ± 0.7	3.4 ± 0.5	3.4 ± 0.5	1.1 ± 0.1
K_1 (M^{-1})	(3.7 ± 3.0) x 10^9	(2.6 ± 0.5) x 10^7	(3.5 ± 0.6) x 10^7	(3.5 ± 0.6) x 10^7	(7.7 ± 0.9) x 10^7
k_2 (s^{-1})		1.0 ± 0.01	1.2 ± 0.03	0.68 ± 0.02	0.39 ± 0.04
k_{-2} (s^{-1})		~0	0.11 ± 0.01	0.19 ± 0.02	0.024 ± 0.023
K_2		-	11 ± 1	3.6 ± 0.4	16 ± 16
k_3 (s^{-1})			0.083 ± 0.015	0.54 ± 0.03	0.66 ± 0.25
k_{-3} (s^{-1})			~0	~0	0.65 ± 0.19
K_3			-	-	1.0 ± 0.5
k_4 (s^{-1})					0.048 ± 0.119
k_{-4} (s^{-1})					~0
K_4					-
χ^2/DOF	4.07941	1.45273	1.0631	1.0631	1.04588
k_d [a]	2.7 x 10^{-3}	~0	~0	~0	~0
$t_{1/2}$ (min)	4.3	-	-	-	-
a	0.53	0.55	0.52	0.59	0.60
b	1.5	0.81	0.62	0.69	0.72
c		1.4	1.4	1.4	2.9
d			1.4	1.5	6.1 x 10^{-4}
e					1.5
bkg	0.32	0.16	0.28	.21	0.20

References:

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst (2010). D66, 213-221 [Doi:10.1107/S0907444909052925] 1–9.

Agarwal, N., and Tyagi, A.K. (2006). Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity. Nucleic Acids Res *34*, 4245–4257.

Artsimovitch, I., Svetlov, V., Murakami, K.S., and Landick, R. (2003a). Co-overexpression of Escherichia coliRNA Polymerase Subunits Allows Isolation and Analysis of Mutant Enzymes Lacking Lineage-specific Sequence Insertions. Journal of Biological Chemistry *278*, 12344–12355.

Bae, B., Chen, J., Davis, E., Leon, K., Darst, S.A., and Campbell, E.A. (2015a). CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. Elife *4*.

Bae, B., Davis, E., Brown, D., Campbell, E.A., Wigneshweraraj, S., and Darst, S.A. (2013). Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of sigma70 domain 1.1. Proc Natl Acad Sci U S A *110*, 19772–19777.

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S.A. (2015b). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. Elife *4*.

Banta, A.B., Chumanov, R.S., Yuan, A.H., Lin, H., Campbell, E.A., Burgess, R.R., and Gourse, R.L. (2013). Key features of sigmaS required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme. Proc Natl Acad Sci U S A *110*, 15955–15960.

Banta, A.B., Cuff, M.E., Lin, H., Myers, A.R., Ross, W., Joachimiak, A., and Gourse, R.L. (2014). Structure of the RNA polymerase assembly factor Crl and identification of its interaction surface with sigma S. J Bacteriol *196*, 3279–3288.

Bao, X., Nickels, B.E., and Fan, H. (2012). Chlamydia trachomatis protein GrgA activates transcription by contacting the nonconserved region of sigma66. Proc Natl Acad Sci U S A *109*, 16870–16875.

Bashyam, M.D., Kaushal, D., Dasgupta, S.K., and Tyagi, A.K. (1996). A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. J Bacteriol *178*, 4847–4853.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature *417*, 141–147.

Bortoluzzi, A., Muskett, F.W., Waters, L.C., Addis, P.W., Rieck, B., Munder, T., Schleier, S., Forti, F., Ghisotti, D., Carr, M.D., et al. (2013). Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. J Biol Chem *288*, 14438–14450.

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., and Heumann, H. (2004). The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. Nat Struct Mol Biol *11*, 551–557.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat Rev Microbiol *2*, 57–65.

Buc, H., and McClure, W.R. (1985). Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry *24*, 2712–2723.

Burgess, R.R. (1969). Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem *244*, 6168–6176.

Burgess, R.R., and Travers, A.A. (1970). Escherichia coli RNA polymerase: purification, subunit structure, and factor requirements. Fed Proc *29*, 1164–1169.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. Nature *221*, 43–46.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell *104*, 901–912.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell *9*, 527–539.

Campbell, E.A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., Severinov, K., and Darst, S.A. (2005). Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. Embo J *24*, 674–682.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep *5*, 1121–1131.

Craig, M.L., Tsodikov, O.V., McQuade, K.L., Schlax, P.E.J., Capp, M.W., Saecker, R.M., and Record, M.T.J. (1998). DNA footprints of the two kinetically significant

intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. J Mol Biol *283*, 741–756.

Czyz, A., Mooney, R.A., Iaconi, A., and Landick, R. (2014). Mycobacterial RNA Polymerase Requires a U-Tract at Intrinsic Terminators and Is Aided by NusG at Suboptimal Terminators. mBio *5*, e00931–14–e00931–14.

Darst, S.A. (2001). Bacterial RNA polymerase. Curr Opin Struct Biol *11*, 155–162.

Davis, C.A., Bingman, C.A., Landick, R., Record, M.T.J., and Saecker, R.M. (2007). Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by Escherichia coli RNA polymerase. Proc Natl Acad Sci U S A *104*, 7833–7838.

Davis, E., Chen, J., Leon, K., Darst, S.A., and Campbell, E.A. (2015). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. Nucleic Acids Res *43*, 433–445.

de Chastellier, C. (2009). The many niches and strategies used by pathogenic mycobacteria for survival within host macrophages. Immunobiology *214*, 526–542.

Dey, A., Verma, A.K., and Chatterji, D. (2010). Role of an RNA polymerase interacting protein, MsRbpA, from Mycobacterium smegmatis in phenotypic tolerance to rifampicin. Microbiology *156*, 873–883.

Dey, A.; Verma, A.K., Chatterji, D. (2011) Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA. Microbiology *157*, 2056–2071.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126–2132.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An allosteric path to transcription termination. Mol Cell *28*, 991–1001.

Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell *147*, 1257–1269.

Feklistov, A., Mekler, V., Jiang, Q., Westblade, L.F., Irschik, H., Jansen, R., Mustaev, A., Darst, S.A., and Ebright, R.H. (2008). Rifamycins do not function by allosteric modulation of binding of Mg2+ to the RNA polymerase active center. Proc Natl Acad Sci U S A *105*, 14820–14825.

Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. Annu. Rev. Microbiol. *68*, 357–376.

Flentie, K., Garner, A.L., and Stallings, C.L. (2016). Mycobacterium tuberculosis

Transcription Machinery: Ready To Respond to Host Attacks. J Bacteriol *198*, 1360–1373.

Forti, F., Mauri, V., Deho, G., and Ghisotti, D. (2011). Isolation of conditional expression mutants in Mycobacterium tuberculosis by transposon mutagenesis. Tuberculosis (Edinb) *91*, 569–578.

Goldman, S.R., Ebright, R.H., and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts in vivo. Science *324*, 927–928.

Gries, T.J., Kontur, W.S., Capp, M.W., Saecker, R.M., and Record, M.T.J. (2010). One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. Proc Natl Acad Sci U S A *107*, 10418–10423.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. *57*, 441–466.

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA Promoter Regulation by Nonoptimal Binding of σ Region 1.2: An Additional Recognition Element for RNA Polymerase. Cell *125*, 1069–1082.

Haugen, S.P., Ross, W., and Gourse, R.L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol *6*, 507–519.

Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.-P., and Brodolin, K. (2012). Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the sigma A-containing RNA polymerase holoenzyme. Nucleic Acids Res *40*, 6547–6557.

Hu, Y., Morichaud, Z., Perumal, A.S., Roquet-Baneres, F., and Brodolin, K. (2014). Mycobacterium RbpA cooperates with the stress-response sigmaB subunit of RNA polymerase in promoter DNA unwinding. Nucleic Acids Res *42*, 10399–10408.

Hubin, E.A., Tabib-Salazar, A., Humphrey, L.J., Flack, J.E., Olinares, P.D.B., Darst, S.A., Campbell, E.A., and Paget, M.S. (2015). Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. Proc Natl Acad Sci U S A *112*, 7171–7176.

Johnson, K.A., Simpson, Z.B., and Blom, T. (2009). Global KinTek explorer: a new computer program for dynamic simulation and fitting of kinetic data. Analytical Biochemistry. Analytical Biochemistry *387*, 20–29.

Ko, J., and Heyduk, T. (2014). Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochem. J. *463*, 135–144.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J Mol Biol *372*, 774–797.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062–1066.

Landick, R., Krek, A., Glickman, M.S., Socci, N.D., and Stallings, C.L. (2014). Genome-Wide Mapping of the Distribution of CarD, RNAP sigma, and RNAP beta on the Chromosome using Chromatin Immunoprecipitation Sequencing. Genom Data *2*, 110–113.

Lane, W.J., and Darst, S.A. (2010a). Molecular evolution of multisubunit RNA polymerases: structural analysis. J Mol Biol *395*, 686–704.

Lane, W.J., and Darst, S.A. (2010b). Molecular evolution of multisubunit RNA polymerases: sequence analysis. J Mol Biol *395*, 671–685.

Leibman, M., and Hochschild, A. (2007). A σ-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. Embo J *26*, 1579–1590.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell *108*, 599–614.

Miropolskaya, N., Ignatov, A., Bass, I., Zhilina, E., Pupov, D., and Kulbachinskiy, A. (2012). Distinct Functions of Regions 1.1 and 1.2 of RNA Polymerase Subunits from Escherichia coli and Thermus aquaticus in Transcription Initiation. Journal of Biological Chemistry *287*, 23779–23789.

Mitchell, J.E., Zheng, D., Busby, S.J.W., and Minchin, S.D. (2003). Identification and analysis of "extended -10" promoters in Escherichia coli. Nucleic Acids Res *31*, 4689–4695.

Mooney, R.A., and Landick, R. (2003). Tethering sigma70 to RNA polymerase reveals high in vivo activity of sigma factors and sigma70-dependent pausing at promoter-distal locations. Genes Dev *17*, 2839–2851.

Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA Polymerase: An On-Again, Off-Again Relationship? Mol Cell *20*, 335–345.

Munson, L.M., and Reznikoff, W.S. (1981). Abortive initiation and long ribonucleic acid synthesis. Biochemistry *20*, 2081–2085.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol *13*, 31–39.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002a). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science *296*, 1280–1284.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002b).

Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science *296*, 1285–1290.

Newell, K.V., Thomas, D.P., Brekasis, D., and Paget, M.S.B. (2006). The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on Streptomyces coelicolor. Mol Microbiol *60*, 687–696.

Newton-Foot, M., and Gey van Pittius, N.C. (2013). The complex architecture of mycobacterial promoters. Tuberculosis (Edinb) *93*, 60–74.

Opalka, N., Brown, J., Lane, W.J., Twist, K.-A.F., Landick, R., Asturias, F.J., and Darst, S.A. (2010). Complete Structural Model of Escherichia coli RNA Polymerase from a Hybrid Approach. PLoS Biol *8*, e1000483.

Osterberg, S., del Peso-Santos, T., and Shingler, V. (2011). Regulation of alternative sigma factor use. Annu. Rev. Microbiol. *65*, 37–55.

Owen, B., and McMurray, C. (2009). Rapid method for measuring DNA binding to protein using fluorescence anisotropy. Protocol Exchange.

Paget, M.S., Molle, V., Cohen, G., Aharonowitz, Y., and Buttner, M.J. (2001). Defining the disulphide stress response in Streptomyces coelicolor A3(2): identification of the sigmaR regulon. Mol Microbiol *42*, 1007–1020.

Pribnow, D. (1975). Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J Mol Biol *99*, 419–443.

Rammohan, J., Ruiz Manzano, A., Garner, A.L., Prusa, J., Stallings, C.L., and Galburt, E.A. (2016). Cooperative stabilization of Mycobacterium tuberculosis rrnAP3 promoter open complexes by RbpA and CarD. Nucleic Acids Res.

Rammohan, J., Ruiz Manzano, A., Garner, A.L., Stallings, C.L., and Galburt, E.A. (2015). CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. Nucleic Acids Res *43*, 3272–3285.

Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., and Gourse, R.L. (1994). Factor Independent Activation of rrnB P1. J Mol Biol *235*, 1421–1435.

Roe, J.H., Burgess, R.R., and Record, M.T.J. (1984). Kinetics and mechanism of the interaction of Escherichia coli RNA polymerase with the lambda PR promoter. J Mol Biol *176*, 495–522.

Ross, W., and Gourse, R.L. (2009). Analysis of RNA polymerase-promoter complex formation. Methods *47*, 13–24.

Ruff, E.F., Drennan, A.C., Capp, M.W., Poulos, M.A., Artsimovitch, I., and Record, M.T.J. (2015). E. coli RNA Polymerase Determinants of Open Complex Lifetime and

Structure. J Mol Biol *427*, 2435–2450.

Rutherford, S.T., Villers, C.L., Lee, J.-H., Ross, W., and Gourse, R.L. (2009). Allosteric control of Escherichia coli rRNA promoter complexes by DksA. Genes Dev *23*, 236–248.

Saecker, R.M., Record, M.T.J., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol *412*, 754–771.

Saecker, R.M., Tsodikov, O.V., McQuade, K.L., Schlax, P.E.J., Capp, M.W., and Record, M.T.J. (2002). Kinetic studies and structural models of the association of E. coli sigma(70) RNA polymerase with the lambdaP(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. J Mol Biol *319*, 649–671.

Schroeder, L.A., and deHaseth, P.L. (2005). Mechanistic Differences in Promoter DNA Melting by Thermus aquaticus and Escherichia coli RNA Polymerases. Journal of Biological Chemistry *280*, 17422–17429.

Schwartz, E.C., Shekhtman, A., Dutta, K., Pratt, M.R., Cowburn, D., Darst, S., and Muir, T.W. (2008). A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. Chem Biol *15*, 1091–1103.

Sclavi, B., Zaychikov, E., Rogozina, A., Walther, F., Buckle, M., and Heumann, H. (2005). Real-time characterization of intermediates in the pathway to open complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. Proc Natl Acad Sci U S A *102*, 4706–4711.

Shimamoto, N., Kamigochi, T., and Utiyama, H. (1986). Release of the sigma subunit of Escherichia coli DNA-dependent RNA polymerase depends mainly on time elapsed after the start of initiation, not on length of product RNA. J Biol Chem *261*, 11859–11865.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007). Anatomy of Escherichia coli 70 promoters. Nucleic Acids Res *35*, 771–788.

Srivastava, D.B., Leon, K., Osmundson, J., Garner, A.L., Weiss, L.A., Westblade, L.F., Glickman, M.S., Landick, R., Darst, S.A., Stallings, C.L., et al. (2013). Structure and function of CarD, an essential mycobacterial transcription factor. Proc Natl Acad Sci U S A *110*, 12619–12624.

Stallings, C.L., Stephanou, N.C., Chu, L., Hochschild, A., Nickels, B.E., and Glickman, M.S. (2009). CarD Is an Essential Regulator of rRNA Transcription Required forMycobacterium tuberculosis Persistence. Cell *138*, 146–159.

Studitsky, V.M., Brodolin, K.L., Liu, Y., and Mirzabekov, A.D. (2001). Topography of lacUV5 initiation complexes. Nucleic Acids Res *29*, 854–861.

Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R.A., Ghosh, S., Parsy, M.-L., Simpson, P.J., O'Dwyer, K., Matthews, S.J., and Paget, M.S. (2013). The actinobacterial

transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. Nucleic Acids Res *41*, 5679–5691.

Tsodikov, O.V., and Record, M.T.J. (1999). General method of analysis of kinetic equations for multistep reversible mechanisms in the single-exponential regime: application to kinetics of open complex formation between Esigma70 RNA polymerase and lambdaP(R) promoter DNA. Biophys J *76*, 1320–1329.

Toutchkine, A., Kraynov, V., and Hahn, K. (2003). Solvent-Sensitive Dyes to Report Protein Conformational Changes in Living Cells. J. Am. Chem. Soc. *125*, 4132–4145.

Tsodikov, O.V., and Record, M.T.J. (1999). General method of analysis of kinetic equations for multistep reversible mechanisms in the single-exponential regime: application to kinetics of open complex formation between Esigma70 RNA polymerase and lambdaP(R) promoter DNA. Biophys J *76*, 1320–1329.

Vassylyev, D.G., Sekine, S.-I., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature *417*, 712–719.

Vo, N.V., Hsu, L.M., Kane, C.M., and Chamberlin, M.J. (2003). In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. Biochemistry *42*, 3798–3811.

Whipple, F.W., and Sonenshein, A.L. (1992). Mechanism of initiation of transcription by Bacillus subtilis RNA polymerase at several promoters. J Mol Biol *223*, 399–414.

Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., van Soolingen, D., Rüsch-Gerdes, S., Locht, C., Brisse, S., et al. (2008). Origin, Spread and Demography of the Mycobacterium tuberculosis Complex. PLoS Pathog *4*, e1000160.

Xue, Y., Hogan, B.P., and Erie, D.A. (2000). Purification and initial characterization of RNA polymerase from Thermus thermophilus strain HB8. Biochemistry *39*, 14356–14362.

Young, B.A., Gruber, T.M., and Gross, C.A. (2004). Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. Science *303*, 1382–1384.

Yuzenkova, Y., Tadigotla, V.R., Severinov, K., and Zenkin, N. (2011). A new basal promoter element recognized by RNA polymerase core enzyme. Embo J *30*, 3766–3775.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811–824.

Zuo, Y., and Steitz, T.A. (2015). Crystal structures of the E. coli transcription initiation complexes with a complete bubble. Mol Cell *58*, 534–540.