Summer 8-15-2021

RNA Polymerase Binding Protein A (RbpA) Regulation of Mycobacteria Transcription and Sensitivity to Fidaxomicin

Jerome Prusa
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Program in Molecular Microbiology and Microbial Pathogenesis

Dissertation Examination Committee:
Christina L. Stallings, Chair
Michael G. Caparon
Eric A. Galburt
Daniel E. Goldberg
Hani S. Zaher

RNA Polymerase Binding Protein A (RbpA) Regulation of Mycobacteria Transcription and Sensitivity to Fidaxomicin

by
Jerome Daniel Prusa

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2021
St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgments

I have many people to thank. First and foremost, thank you to my scientific mentors. I am forever grateful for my thesis mentor Christina Stallings who more than anyone has shaped my development as a scientist through countless hours of invaluable mentorship and support. In addition to her individual contributions, I am thankful for the lab Christina has built, which has consistently been filled with outstanding people that I have enjoyed working with and learning from. I would particularly like to thank Jeremy Huhyn, Ashley Garner, Katherine Mann, Dennis Zhu, Greg Harrison, Michael Nehls, Helen Blaine, Skyler Hendrix and Gustavo Santiago-Collazo for their mentorship, advice, ideas and friendship both in and outside the lab. I am grateful for William Tapprich and Paul Davis at the University of Nebraska-Omaha, who sparked my curiosity and passion for research. I have benefited from and appreciate being a member of the Washington University School of Medicine Department of Microbiology that is well-managed by excellent administrative, academic and technical staff, who I owe many thanks. Finally, I would like to thank my friends and family, especially my partner Reyka Jayasinghe, my parents Dan and Mary Prusa, my siblings Jessica and Alex Prusa, my brother-in-law Nelson Flores and sister-in-law Liza Martinez. The importance of your love and support to me cannot be overstated.

Jerome Prusa
Washington University in St. Louis
August 2021

Dedication:
To my parents Daniel and Mary Prusa for their love and support.

ABSTRACT OF THE DISSERTATION

RNA Polymerase Binding Protein A (RbpA) Regulation of Mycobacteria Transcription and Sensitivity to Fidaxomicin
by

Jerome Daniel Prusa

Doctor of Philosophy in Biology and Biomedical Sciences

Program in Molecular Microbiology and Microbial Pathogenesis

Washington University in St. Louis, 2021

Professor Christina L. Stallings Ph.D, Chair

Mycobacterium tuberculosis is the causative agent of the disease tuberculosis (TB) and remains one of the deadliest microorganisms on the planet. The effort to eradicate *M. tuberculosis* would benefit from the development of novel therapeutics, which requires a detailed understanding of *M. tuberculosis* physiology. Like all living organisms, *M. tuberculosis* gene expression requires transcription. Transcription in the phylum Actinobacteria, which includes mycobacteria, is unique because it includes RNA Polymerase Binding Protein A (RbpA) that is essential in both *M. tuberculosis* and the nonpathogenic model organism *Mycobacterium smegmatis*.

RbpA increases the housekeeping σ^A and housekeeping like σ^B interactions with the RNA polymerase (RNAP) and can increase transcription by both σ^A and σ^B bound RNAPs *in vitro*, suggesting that RbpA activates *M. tuberculosis* transcription. During transcription initiation, the equilibrium between the melted and unmelted promoter conformations is a common regulatory target. RbpA stabilizes the melted DNA conformation called the RNA polymerase open promoter complex (RP_o). Structural studies revealed that RbpA is comprised of four structural domains including the N-terminal tail (NTT), core domain (CD), basic linker (BL) and sigma interaction domain (SID). RbpA BL interacts with the DNA phosphate backbone of the non-template strand while the SID mediates RbpA's interaction with σ^A and σ^B. The activities of both the BL and SID are important for RbpA RP_o stabilizing activity *in vitro*. Using a panel of RbpA point mutants and RbpA domain truncation mutants, I further characterized the activities of RbpA's four structural domains *in vitro* and *in vivo*. The activities of all four domains are required for *M. tuberculosis* growth while only the BL and SID are required for *M. smegmatis* growth. RNA-sequencing analysis revealed that RbpA activates transcription of some genes while repressing the transcription of other genes, and the activities of the BL/SID and NTT/CD affect transcription of two distinct gene subsets. We determined that the SID is necessary and sufficient for RbpA interaction with both the σ^A and σ^B bound RNAPs and weakening RbpA's interaction with the RNAP decreases RbpA protein levels in *M. smegmatis*. *In vitro* analysis done in collaboration with the Galburt lab revealed that the BL and SID are required for RbpA's RP_o stabilizing activity while the NTT and CD antagonize RbpA RP_o stabilizing activity.

Structural studies show that the NTT and CD are positioned near multiple RNAP-σ^A holoenzyme functional domains, suggesting that the RbpA NTT and CD could have a number of

effects on RNAP activity. However, these studies did not identify which contacts between the NTT or CD and the RNAP mediate the antagonism of RP_o stability that we observed in our studies. In addition, structural studies predict that the RbpA NTT contributes contacts to the binding site for the antibiotic fidaxomicin (Fdx) on the RNAP. Deletion of the NTT results in a decrease in *M. smegmatis* sensitivity to Fdx, but whether this is caused by a loss of contacts with Fdx was unknown. Using a panel of *rbpA* mutants with single amino acid substitutions replacing conserved residues within the NTT, I probed what RbpA NTT residues are involved in regulating Fdx activity and RP_o stability. We identify multiple residues in the NTT along with other RbpA domains that contribute to Fdx activity *in vivo*. We also identify RbpA NTT residues that contribute to antagonism of RbpA-mediated stabilization of RP_o and link this antagonism to increased full length transcript production.

In work characterizing the role of RbpA's interaction with σ^B I determined that the loss of RbpA BL or SID activities alters *sigB* from its typical status as a non-essential gene to a synthetically essential gene. RNA-sequencing analysis of *M. smegmatis* with a *sigB* deletion (Δ*sigB*) shows that *sigB* regulates a cohort of transcripts that if translated encode short and highly charged proteins. In addition, the subset of transcripts differentially expressed in *M. smegmatis* Δ*sigB* shares little overlap with the gene subset differentially expressed in *M. smegmatis* expressing *rbpA* with a point mutation in the SID that weakens RbpA interaction with both σ^A and σ^B, suggesting that RbpA-independent σ^B regulation occurs during logarithmic growth.

My thesis work has improved our understanding of RbpA regulation of mycobacteria transcription and RbpA's role in fidaxomicin activity. This work shows that RbpA regulates transcription through novel mechanisms, shedding new light on the similarities and differences

between the Actinobacteria and *E. coli* paradigms of bacterial transcription. Furthermore, the design of future therapeutics might benefit from this interrogation of RbpA activities and how this essential protein contributes to Fdx activity against mycobacteria.

Chapter 1: Introduction – Overview of Bacterial Transcription

Jerome Prusa, Christina L. Stallings

The Bacterial RNA Polymerase

For all cellular organisms, transcription is carried out by a DNA dependent RNA polymerase (RNAP) that links nucleotide triphosphates together through a phosphodiester bond [1]. The structure of archaeal, bacterial and eukaryotic RNAPs are similar in resembling a crab claw with two mobile pincer domains that open and close around a region called the primary channel [2–4]. At the base of the primary channel is Mg^{2+} required for catalysis of the phosphodiester bond. In addition to the primary channel, which accommodates DNA, RNAPs of all living organisms contain two additional channels including the secondary channel where substrate NTPs enter the RNAP, and the RNA exit channel where an elongating RNA transcript leaves the RNAP.

The core bacterial RNAP (bRNAP) consists of five protein subunits including two identical α subunits, β , β' and ω, discovered by Richard Burgess when he purified the *E. coli* core RNAP **(Figure 1)**[5]. The bRNAP is a combination of highly conserved regions that are shared among all bRNAPs and regions that are unique to the RNAPs of certain subsets of bacteria [6,7]. Most of the highly conserved regions of the RNAP are located within the interior of the core RNAP and most of these regions are required for either catalysis of phosphodiester bond formation or interactions with DNA or RNA. These conserved bRNAP core regions among others include the β' clamp that makes up one pincer of the crab claw, the β protrusion and β lobe that together make up the second pincer of the crab claw [8–10], the bridge helix and trigger loop that mediate NTP addition to the growing RNA [11,12], and the base of the primary channel containing the NADFDGD motif that constitutes the active center of the RNAP [13,14]. The highly conserved regions of the RNAP are connected to one another by lineage specific insertions that

in some cases are hundreds of amino acids in length and generally reside on the exterior of the bacterial RNAP [6,7]. The function of these lineage specific insertions is mostly unknown.

In addition to the core RNAP all bacteria have at least one housekeeping σ factor that binds to the RNAP to form the holoenzyme and is responsible for the expression of essential genes [15]. The number of σ factors encoded by a bacterium varies widely from a single σ factor such as for *Mycoplasma* to *Plesiocytis pacifica*, which is the current record holder, encoding at least 118 σ factors **(collected from MiST 3.0 https://mistdb.com)** [16]. There are two groups of σ factors in bacteria including the σ^{70}-like group and the σ^{54}-like group. The majority of bacterial transcription research has focused on the σ^{70}-like group, which are able to melt DNA in an ATP-independent manner unlike the σ^{54}-like group that require ATP for DNA melting [17]. Despite being functionally homologous with both σ factor groups being capable of initiating transcription, the σ^{70}-like and σ^{54}-like groups are structurally distinct from one another and initiate transcription through distinct mechanisms. *M. tuberculosis* encodes 13 σ^{70}-like σ factors and zero σ^{54}-like σ factors and therefore from this point on any mention of σ factors is in reference to σ^{70}-like σ factors [18].

Structurally, σ factors are comprised of a variable number of globular domains that are connected to one another by flexible linker domains. There are four types of σ^{70}-like σ factors organized by domain complexity (**Figure 2)** [19]. Type I σ factors are the most complex and include three highly conserved globular domains, domains 2, 3, and 4, an unstructured and poorly conserved N-terminal domain 1.1 and a non-conserved domain that connects region 1.1 and region 2. Type II σ factors are structurally similar to type I σ factors except the type II σ factors lack domain 1.1 [20]. Type III σ factors lack both domain 1.1 and the non-conserved

domain, while type IV σ factors are comprised of only domains 2 and 4 [20,21]. In addition to these five domains used to categorize σ^{70}-like σ factors into types I - IV, type I σ factors' domains 2 and 3 are further divided into domains 2.1 – 2.4 and domains 3.0 – 3.2, while both type I and type II σ factors have an additional region 1.2 located between domain 1.1 and the non-conserved domain in the type I σ factors or located at the N-terminus of type II σ factors (**Figure 2**). The conserved globular domains 2, 3 and 4 are important for allowing σ to bind the RNAP or DNA in a similar manner across bacterial species [8–10]. The function of the non-conserved domain is not yet known but it may provide species specific binding sites for transcription factors [22]. The functions performed by region 1.1 continues to grow. The functions of region 1.1 include preventing free σ binding to DNA by interacting with the σ C-terminus, which maintains a compact σ structure that masks σ's DNA binding domains [23,24], preventing non-specific holoenzyme interaction with DNA by occupying the holoenzyme primary channel as a DNA-mimic [25,26] and suppressing *E. coli* rRNA expression during the stringent response [27]. Region 3.2 is a conserved unstructured linker in type I σ factors that connects the globular portion of domain 3 and domain 4. Like region 1.1, many different functions have been described for region 3.2 including NTP substrate binding via region 3.2 interactions with the DNA template strand [28,29], regulating the transition from transcription initiation to transcription elongation [29,30] and contributing to *E. coli*'s stringent response [31].

Upon core RNAP binding, σ undergoes a conformational change that repositions σ to span the length of the RNAP, allowing multiple interactions between σ and the core RNAP and correctly positions σ's DNA binding regions for sequence specific DNA interactions [32–34]. The extra domains in type I housekeeping σ^{70} mediates more interactions between type I σ factors

and core RNAP, which likely explains why type I σ factors have the highest affinity for core RNAP [35]. However, type I - IV σ factors all bind to conserved core RNAP structures in a similar manner that involves σ region 2 binding to the RNAP β' coiled coil and σ region 4 binding to the β flap tip helix [36–40].

The Steps of Bacterial Transcription

Bacterial Promoter Composition and RNAP Holoenzyme Promoter Recognition

The holoenzyme interacts with DNA sequence elements that are collectively termed the promoter. Bacteria holoenzymes utilize at least six DNA promoter sequence elements to bind to DNA and initiate transcription. The six promoter sequence elements that have been discovered thus far include the -10 element, -35 element, extended -10 element, discriminator region, UP element and the core recognition element (CRE) **(Figure 3)**.

The -10 element was the first promoter sequence element discovered independently by both Pribnow and Schaller et al [41,42]. The consensus -10 element in *E. coli* is 5' – TATAAT – 3' and most commonly occupies positions -12 to -7 relative to the +1 transcription start site (TSS). A highly conserved group of aromatic residues in region 2 of σ^{70} interact with the -10 element [8,19].

The -35 element was discovered shortly after the -10 element and was determined to have the consensus sequence 5' – TTGACA – 3' in *E. coli* [43,44]. Initial holoenzyme recognition of the promoter occurs at the upstream sequence elements including the -35 element [45]. A helix-turn-helix motif in σ^{70} region 4 fits into the minor groove of the -35 element, which stabilizes the RNAP-DNA interaction and positions the holoenzyme for subsequent interactions with downstream promoter sequence elements [46].

The discriminator region was the third sequence element to be discovered through characterization of promoters regulated by (p)ppGpp during the stringent response in *E. coli* [47]. The 'stringent' promoters or those that are repressed during the stringent response were found to have an enrichment of cytosine at positions -5 to +1 but a consensus discriminator sequence was not defined. Analyses almost two and half decades after the initial description of the discriminator region defined the consensus discriminator region in *E. coli* as 5' – GGGA – 3' at positions -6 to -3 on the non-template DNA strand [48,49]. The discriminator region interacts with σ^{70} region 1.2 [8].

The extended -10 element is located at positions -15 and -14 directly upstream of the -10 element. In *E. coli,* the consensus extended -10 element is -15T and -14G. The extended -10 element was identified through the discovery that certain sequences immediately upstream of the -10 element could compensate for loss of holoenzyme activity in the absence of the -35 element [50,51]. Regions 3.0 and 3.1 of σ^{70} interact with the extended -10 element [52,53].

The UP element is an AT rich region upstream of the -35 element, typically located between positions -40 and -60. The UP element is recognized by the C-terminal domains of the RNAP α subunits rather than through σ, which mediates most of the holoenzyme's interactions with promoter sequence elements [54]. In addition to σ^{70} region 4 binding the -35 element, the RNAP α interactions with the UP element are the first interactions between the holoenzyme and DNA [55].

The CRE is a guanosine at the +2 position of the non-template strand that interacts with RNAP β residues that form a base specific binding pocket for guanosine [8]. First mention of what was later determined to be the CRE was observed when the portion of RNAP β that forms the CRE binding pocket was found to cross-link to positions -6 - +2 of the promoter [56]. Later

structural studies revealed why the crosslinking occurred by resolving a promoter bound holoenzyme structure with the +2 guanosine sequestered in a base specific binding pocket [8].

These six promoter sequence elements as a collective create many opportunities for the holoenzyme to recognize and interact with any given promoter. This flexibility in what can constitute a functional promoter results in a rich diversity of promoter architectures found within the bacterial genome.

Promoter Melting

The holoenzyme stably bound to the DNA promoter prior to promoter melting is called the RNAP closed promoter complex (RP_c). The holoenzyme then unwinds up to 13-bp of DNA from -11 to +2 to form a complex called the RNAP open promoter complex (RP_o). The transition from RP_c to RP_o is collectively called isomerization and has been an area of active investigation for nearly 60 years. Our current understanding of isomerization was born out of extensive biochemical and genetic dissection of *E. coli* σ^{70} holoenzyme activity on a handful of model promoters, which established that isomerization is a multi-step process [45,57–62]. Understanding of isomerization is now being refined through the combination of technological breakthroughs in structural biology that are allowing isomerization intermediates to be captured and studies that compare isomerization in *E. coli* with other bacterial species [63–70].

Isomerization begins with a group of highly conserved aromatic residues in σ region 2.3 unstacking -11A through a base flipping mechanism [71,72]. After unstacking -11A, residues in σ region 2.3 sequester -11A in a binding pocket [73]. The conserved amino acids that constitute the -11A binding pocket make adenosine specific interactions allowing the -11 position of the -10 element to be read [73]. The σ region 2.3 adenosine specific binding pocket and the importance of this base in nucleating promoter melting explains why the -11A is the most conserved position

not only among the six positions constituting the -10 element but across all the promoter sequence elements utilized by housekeeping σ bound holoenzyme in both *E. coli* and *M. tuberculosis*. Early promoter melting extends from position -11 to -8 with the DNA located outside of the holoenzyme primary cleft and the DNA clamp in the closed conformation [70]. Propagation of promoter melting from the -8 to -4 position is thought to coincide with the removal of σ^{70} region 1.1, which acts as a DNA mimic, from the holoenzyme primary cleft [70]. Ejection of σ^{70} region 1.1 from the primary channel now allows positions -8 to +2 of the ssDNA template strand to enter and occupy the primary channel [68,70]. During this second phase of promoter melting -7T is sequestered in a base specific manner by a second σ region 2.3 nucleotide binding pocket [73]. The invasion of DNA positions -8 to +2 of the ssDNA template strand into the primary channel likely requires temporary RNAP clamp opening before it closes again and then remains closed to the completion of isomerization [68,70,74]. The final step is the melting from positions -4 to +2, which is paired with increased interactions between the RNAP clamp and downstream dsDNA, finally resulting in RP_o [70,75,76]. In addition to the -11A and -7T positions being sequestered in binding pockets formed by σ region 3.2, +2 can also be sequestered in a binding pocket formed by residues in β, which all together stabilize RP_o [8].

Promoter Escape

Once stable RP_o has formed, the holoenzyme begins the process of promoter escape. During promoter escape a nascent transcript is formed while the holoenzyme maintains the contacts that are formed in RP_o. The growth of the nascent transcript requires downstream DNA to be pulled into the RNAP resulting in a process called 'scrunching' [77,78]. As more downstream DNA is pulled into the holoenzyme primary channel while the nascent RNA grows, the contacts between the RNAP and the promoter are strained [79]. The increasing strain is relieved through one

of two ways. Either the contacts between the RNAP and promoter are broken allowing the RNAP to escape and progress forward to elongation or the nascent RNA transcript is released, the RNAP repositions at the transcription start site and the process restarts [80,81]. The RNA that is released prematurely is called an abortive transcript and at some promoters several abortive transcripts will be produced [82]. When the RNA transcript elongates to 5-6 nucleotides in length, the growing transcript collides with σ region 3.2, a key event during RNAP's 'decision' to progress to elongation or produce an abortive transcript [28,34]. Region 3.2 of σ blocks the RNA exit channel that the 5' end of the transcript approaches as it grows [67,83]. Therefore, the transcript must displace σ region 3.2 in order to continue growing towards the RNA exit channel. If the transcript displaces σ region 3.2, a conformational change in σ occurs that initiates the breaking of σ-RNAP interactions and allows σ dissociation from core RNAP for elongation.

Elongation and Termination

The focus of this thesis is transcription initiation and therefore I have decided omit an in depth overview of transcription elongation and termination. However, both transcriptional elongation and termination in bacteria include an ever-growing number of important features and our current understanding of both topics have been reviewed elsewhere [84–87].

Mycobacteria Transcription

Unique Features of the Mycobacteria RNAP and Mycobacteria Promoter

The description of bacterial transcription included above is mostly derived from studying *E. coli*. This description is useful as a starting point, but a more complete understanding of bacterial transcription requires interrogating these processes in more than one species. Characterization of transcription in mycobacteria and identifying similarities and differences when compared to *E. coli* is broadening our understanding of transcription. By identifying

similarities between *E. coli* and mycobacteria transcription we gain insight into what features are similar across bacteria. Conversely, identifying differences better defines the Actinobacteria paradigm of bacterial transcription to complement the better understood γ-proteobacteria paradigm and points to 'customizable' features of transcription that can evolve to meet the needs of different bacterial species.

Unique features of mycobacteria transcription include lineage specific insertions in the RNAP β' subunit and σ^A, unique promoter architecture and an Actinobacteria specific combination of two essential proteins that interact with the RNAP **(Figure 4 and 5)**. The mycobacteria core RNAP subunit composition is like most bacteria and includes the β and β' subunits, two α subunits and the ω subunit. However, the mycobacteria β' subunit harbors a lineage specific 90 amino acid insertion conserved in the Actinobacteria phylum [6]. Another insertion occurs in mycobacteria's housekeeping σ^A, where there is a 163 amino acid tail attached the N-terminus giving mycobacteria an unusually long σ^A domain 1.1. The 90 amino acid β' insertion forms two anti-parallel α helices that protrude from the downstream facing RNAP clamp domains [10]. The function of the β' insertion is unknown; however, the insertion is positioned such that it blocks the primary channel of the RNAP, hinting that this structure may play a role in regulating DNA's entrance into the primary cleft during isomerization [10]. The σ^A domain 1.1 is predicted to be unstructured and has evaded resolution in all the mycobacteria RNAP structures published [10,67,68,88]. *E. coli* σ^{70} includes an N-terminal domain 1.1 that is shorter than its counterpart in mycobacteria and is known to occlude DNA from entering the primary cleft during transcription initiation among other functions. It is not yet clear whether mycobacteria's σ^A domain 1.1 performs the same roles as *E. coli*'s σ^{70} domain 1.1.

In addition to the σ^A N-terminal domain 1.1, another unique σ-related feature of mycobacteria transcription is the cohort of alternative σ factors that with σ^A collectively make up the mycobacteria sigma factor network. The number of σ factors encoded in different mycobacterial species varies across a spectrum with *M. smegmatis* encoding a total of 26 σ factors and other species such as *M. tuberculosis* encoding only half as many. The *M. tuberculosis* σ factor network and the role of each σ factor within this network is the best characterized among the mycobacteria species. The 13 σ factors encoded in *M. tuberculosis* include one type I σ factor (σ^A), one type II σ factor (σ^B), one type III σ factor (σ^F) and 10 type IV σ factors (σ^C, σ^D, σ^E, σ^G, σ^H, σ^I, σ^J, σ^K, σ^L, σ^M) [18,89]. Only the housekeeping σ^A is essential in mycobacteria. The 12 non-essential σ factors collectively mediate *M. tuberculosis* adaptation to various stresses that the bacterium encounters in the human host. An important feature of *M. tuberculosis*'s σ factors is the complexity of the regulatory network that often includes multiple σ factors working together to mount an appropriate stress response. The stresses each non-essential *M. tuberculosis* σ factor is important for has been reviewed elsewhere and include stationary phase, starvation, pH stress, high temperature, hypoxia, oxidative stress, surface stress, DNA damage, iron stress, antibiotics and interactions with macrophages [18,89].

Multiple analyses of mycobacteria promoters have shown that mycobacteria promoters differ from *E. coli* promoters in several ways and indicate that the mycobacteria promoter architecture is less stringently defined compared to that of *E. coli* [10,90–92]. The consensus -10 element in both bacteria includes -11A and -7T; however, aside from these two positions the identity of the other 4 positions (-12, -10, -9 and -8) is not defined in mycobacteria. The number of *M. tuberculosis* promoters that include the consensus -35 element and the identity of the consensus *M. tuberculosis* -35 element has varied depending on the criteria used to define the -35

element [10,90–92]. With that said, even when the most liberal criteria are applied to determining whether a *M. tuberculosis* promoter contains a -35 element, the frequency at which -35 elements are found in *M. tuberculosis* promoters is half the frequency of that in *E. coli* [10]. The UP element defined by tracts of AT-rich sequences bound by the C-terminal domain of the RNAP α subunits is also found less frequently in mycobacteria promoters compared to *E. coli* promoters [10], which is not surprising given the large difference in GC content between the genomes. Mycobacteria promoters include the extended -10 element at a frequency comparable to *E. coli* [10,90].

The cohort of transcription factors encoded in *E. coli* and mycobacteria differ. Two well characterized transcription factors, DksA and Fis, are notably absent in mycobacteria. DksA is a secondary channel binding protein important for *E. coli*'s stringent response and Fis is a nucleoid associated protein (NAP) important for organizing the architecture of the bacterial chromosome. Conversely, mycobacteria encode factors that *E. coli* lacks. Two of these factors are CarD and RNA polymerase binding protein A (RbpA).

CarD

Mycobacteria CarD was discovered by Christina Stallings in a screen identifying factors important for *M. tuberculosis*'s DNA damage response and is conserved in several bacterial phyla including Actinobacteria [93]. CarD is comprised of two structural domains including the N-terminal RNAP interaction domain (RID) and the C-terminal DNA binding domain (DBD)[94,95]. CarD's RID binds to the RNAP β lobe on the downstream facing side of the RNAP while CarD's DBD binds DNA non-specifically and through the activity of W85 that acts as wedge at the upstream edge of the transcriptional bubble in RP_o [94,96,97]. The activities of both CarD's domains are required for mycobacteria growth and viability, tolerance to clinically relevant antibiotics, expression of rRNA and *M. tuberculosis* replication in murine lung and spleen tissue

[94,95,98]. During logarithmic growth CarD activates and represses two distinct gene subsets that each constitute roughly one third of *M. tuberculosis*'s genes [99]. ChIP-seq analysis of CarD localization in *M. smegmatis* showed that CarD is present at all promoters [100]. These data show that CarD plays an important role regulating most if not all of *M. tuberculosis*'s gene expression. The mechanism by which CarD activates transcription at one set of promoters while repressing transcription at a second set of promoters is an active area of investigation in Eric Galburt's lab and ours. Briefly, data so far points to a model where CarD increases RP_o stability which in turn decreases promoter escape. Promoters with intrinsically high RP_o stability and low rates of promoter escape are repressed by CarD while promoters with intrinsically low RP_o and high promoter escape rates are activated [64,65,69,99].

RbpA

RbpA was discovered in a screen identifying genes upregulated in *Streptomyces coelicolor* during disulfide stress [101]. Many different types of stress including oxidative stress, stationary phase, starvation, high temperature, antibiotics and interactions with macrophages increase *rbpA* levels in *M. tuberculosis* [102–105]. RbpA is conserved only among Actinobacteria and is essential in mycobacteria. RbpA is comprised of four structural domains including the N-terminal tail (NTT), core domain (CD), basic linker (BL) and sigma interaction domain (SID) (Figure 4). Structures of RbpA bound RP_o show that RbpA extends the length of the RNAP and interacts with several structural regions of the RNAP **(Figure 5)**. The NTT interacts with or is positioned near the β switch 3 region, β' lid, σ region 3.2 and the β' zinc binding domain (ZBD) [67]. RbpA CD interacts with β' ZBD and β' zipper [67,106]. RbpA BL interacts with the DNA non-template strand phosphate backbone at positions -13 and -14 upstream of the -10 element and RbpA SID interacts with the non-conserved region of σ as well as regions 1.2 and 2.3 [22]. RbpA

forms a stable binary complex with *M. tuberculosis* σ^A and σ^B and increases σ^A and σ^B binding to core RNAP [107,108]. *In vitro*, RbpA increases transcription by σ^A and σ^B holoenzyme.

The combined effect of RbpA and CarD on transcription initiation has been interrogated *in vitro* [65,69]. Together RbpA and CarD increase σ^A-bound *M. tuberculosis* RP_o at the *M. tuberculosis* ribosomal RNA AP3 (*rrn*AP3) promoter to a level comparable to *E. coli* RNAP-σ^{70} holoenzyme and to a high level of RP_o than either factor can achieve individually [65]. However, CarD increases *M. tuberculosis* RNAP-σ^A RP_o stability more than RbpA. Detailed kinetic analysis of CarD's effect on RP_o stability at different concentrations revealed that CarD increases RP_o through a two-tier mechanism [64]. At low concentrations, CarD binds *M. tuberculosis* RNAP-σ^A RP_o and prevents transcription bubble collapse and at higher concentrations in excess of *M. tuberculosis* RNAP-σ^A RP_o, CarD will bind *M. tuberculosis* RNAP-σ^A RP_c and promote DNA melting to increase RP_o. The comparison of *M. tuberculosis* RNAP-σ^A and *E. coli* RNAP-σ^{70} RP_o stability on the *M. tuberculosis rrn*AP3 highlighted an important difference in transcription initiation between the two bacterial species, highlighting the intrinsically higher *E. coli* RNAP-σ^{70} RP_o stabilizing activity. During promoter escape, both RbpA and CarD individually decrease the rate of *M. tuberculosis* RNAP-σ^A promoter escape [69]. Similar to RP_o stabilizing activity, CarD has a greater impact on slowing *M. tuberculosis* RNAP-σ^A promoter escape. In the presence of both factors, the rate of promoter escape is increased compared to in the presence of CarD alone, indicating that RbpA can increase the rate of promoter escape in the context of CarD. These *in vitro* analyses together indicate that the combined activity of RbpA and CarD is required for regulating complex transcription initiation kinetics in *M. tuberculosis*.

Figures

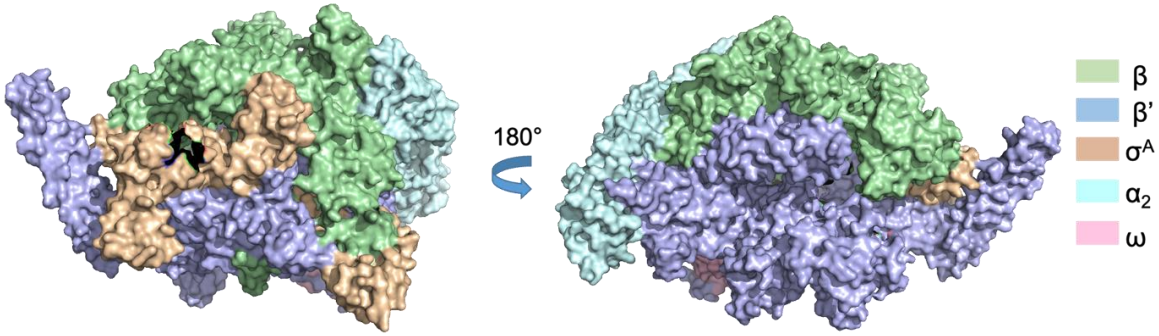


Figure 1. *Mycobacterium tuberculosis* RNA Polymerase RNAP-σ^A holoenzyme. PDB: 6C05 [67]

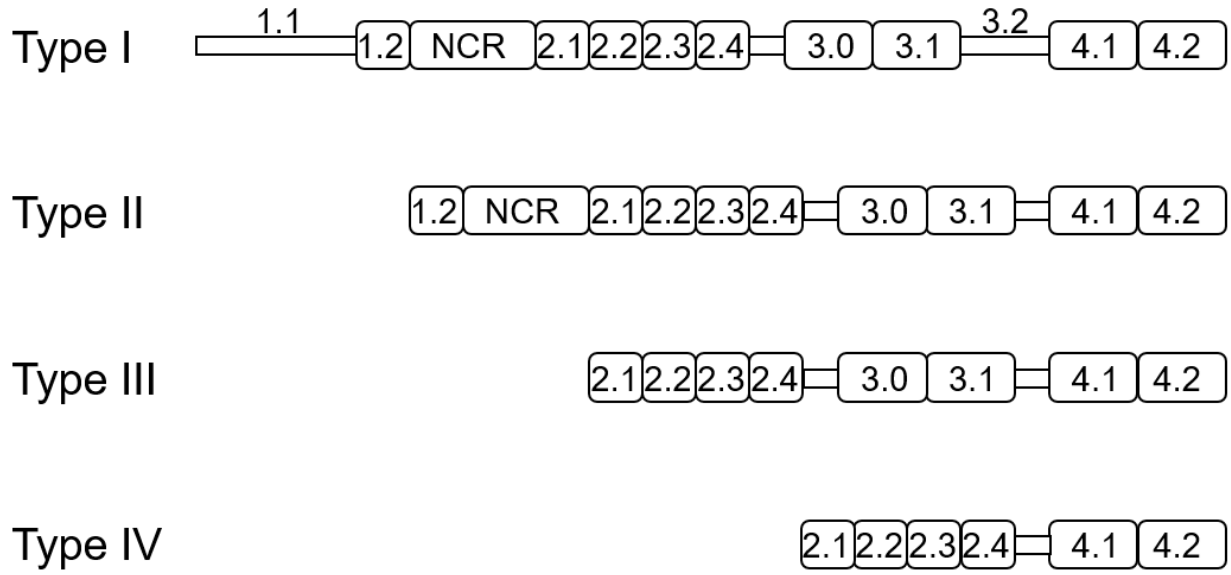


Figure 2. Type I – IV Bacterial σ^{70}-like domain architecture.

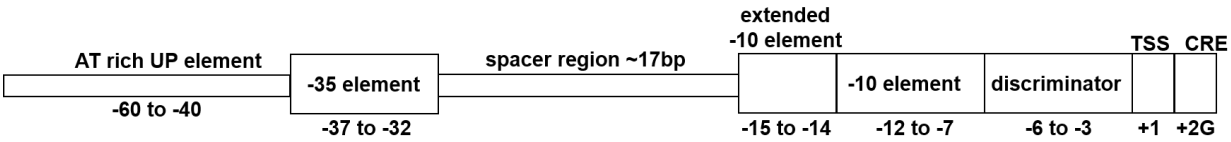


Figure 3. Bacterial Promoter Sequence Elements.

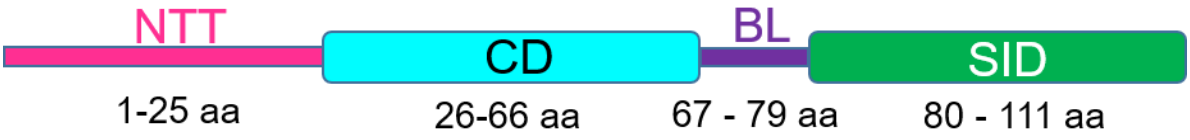


Figure 4. *M. tuberculosis* RNA Polymerase Binding Protein A (RbpA) domain architecture.

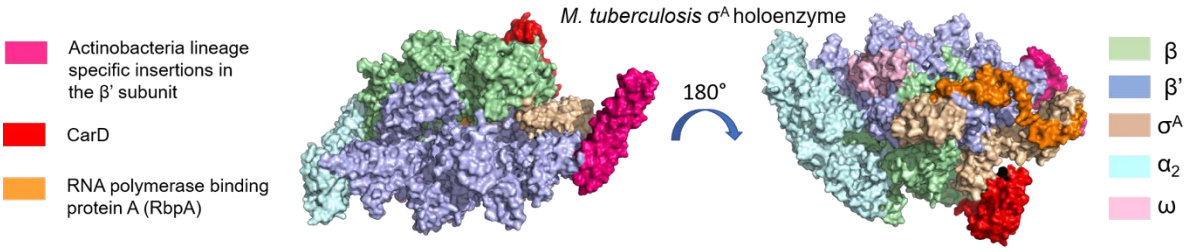


Figure 5. Structure of RbpA and CarD bound *M. tuberculosis* RNAP-σ^A holoenzyme with Actinobacteria lineage specific insertion highlighted. PDB: 6C05 overlayed with 4XLR [67,96]

Chapter 2: Domains within RbpA Serve Specific Functional Roles That Regulate the Expression of Distinct Mycobacterial Gene Subsets

Jerome Prusa, Drake Jensen, Gustavo Santiago-Collazo, Steven S. Pope, Ashley L. Garner, Justin J. Miller, Ana Ruiz Manzano, Eric A. Galburt, Christina L. Stallings

A version of this chapter was originally published as **Prusa et al. 2018 J. Bacteriology**.

Abstract

The RNA polymerase (RNAP) binding protein A (RbpA) contributes to the formation of stable RNAP-promoter open complexes (RPo) and is essential for viability in mycobacteria. Four domains have been identified in the RbpA protein, i.e., an N-terminal tail (NTT) that interacts with RNAP β' and σ subunits, a core domain (CD) that contacts the RNAP β' subunit, a basic linker (BL) that binds DNA, and a σ-interaction domain (SID) that binds group I and group II factors. Limited in vivo studies have been performed in mycobacteria, however, and how individual structural domains of RbpA contribute to RbpA function and mycobacterial gene expression remains mostly unknown. We investigated the roles of the RbpA structural domains in mycobacteria using a panel of rbpA mutants that target individual RbpA domains. The function of each RbpA domain was required for Mycobacterium tuberculosis viability and optimal growth in Mycobacterium smegmatis. We determined that the RbpA SID is both necessary and sufficient for RbpA interaction with the RNAP, indicating that the primary functions of the NTT and CD are not solely association with the RNAP. We show that the RbpA BL and SID are required for RP_o stabilization in vitro, while the NTT and CD antagonize this activity. Finally, RNA-sequencing analyses suggest that the NTT and CD broadly activate gene expression, whereas the BL and SID activate or repress gene expression in a gene dependent manner for a subset of mycobacterial genes. Our findings highlight specific outcomes for the activities of the individual functional domains in RbpA.

Introduction

RbpA was discovered in *Streptomyces coelicolor* as a protein that coimmunoprecipitates with the RNAP and is unique to the *Actinobacteria* phylum [101]. RbpA consists of a central core domain (CD) flanked by an unstructured 26-amino-acid N-terminal tail (NTT) and a C-terminal -

interaction domain (SID) linked to the CD by a 15-amino-acid basic linker (BL) (**Chapter 1, Figure 4**)[22,101,109,110]. The RbpA SID forms a stable binary complex with group I (σ^A in *M. tuberculosis*) and certain group II (σ^B in *M. tuberculosis*) factors [22,101,106,110]. Bacterial two-hybrid experiments in *S. coelicolor* showed that mutating the R88 residue within the RbpA SID to an alanine significantly weakened the interaction between *S. coelicolor* RbpA and the housekeeping HrdB [110], highlighting the importance of this residue in the interaction. Based on structural studies, the BL makes electrostatic contacts with the DNA phosphate backbone of the nontemplate strand upstream of the -10 promoter element in the RP_o conformation, the CD is positioned near the RNAP β' zinc binding domain, and the NTT threads into the RNAP active site cleft between the β' zinc binding domain and the $\sigma^A{}_4$ domain [22,67,106]. In support of a functional role for the BL, fluorescence anisotropy experiments showed that addition of *M. tuberculosis* RbpA to *Mycobacterium bovis* RNAP-σ^A holoenzyme in the presence of *M. tuberculosis* CarD decreased the dissociation constant (K_d) of RNAP binding to a *vapB10* promoter template and an R79A mutation in the *M. tuberculosis* RbpA BL abolished the RbpA mediated increases in RNAP affinity for the *vapB10* promoter [22]. Most characterization of RbpA has been performed *in vitro*, and there have been only limited studies of how the domains of RbpA contribute to gene regulation in mycobacteria. In a recent study using *Mycobacterium smegmatis*, an R79A mutation in the RbpA BL and deletion of the NTT and CD resulted in slower growth of the bacteria [106]. Herein, we expand on that work and compare the roles of each RbpA domain, both *in vitro* and *in vivo*, to show that only the SID is required for association with the RNAP and the activities of different domains affect the expression of distinct gene sets in the bacteria.

Experimental Procedures

Media and bacterial strains. (i) *Mycobacterium tuberculosis*. The Erdman strain was grown at 37°C in 7H9 (broth) or 7H10 (agar) medium supplemented with 60 μl/liter oleic acid, 5 g/liter bovine serum albumin (BSA), 2 g/liter dextrose, and 0.003 g/liter catalase (oleic acid-albumin-dextrose-catalase [OADC]), 0.5% glycerol, and 0.05% Tween 80 (broth). The *M. tuberculosis* merodiploid strain was constructed by integrating pMSG430-*rbpA*$_{Mtb}^{WT}$ (expressing RbpA$_{Mtb}^{WT}$ from a constitutive P*myc1-tetO* promoter; kanamycin resistant) into the *attB* site of the Erdman strain. A specialized transducing phage with homology to *M. tuberculosis* H37Rv nucleotides 2307223 to 2307826 and 2303122 to 2308681 was used to replace all except the start and stop codons of the endogenous *rbpA* gene with a hygromycin resistance cassette in the merodiploid strain, thus generating *ΔrbpA attB*::tet-*rbpA*$_{Mtb}^{WT}$. Gene swapping was used to construct strains of mycobacteria expressing different *rbpA* alleles and to test their viability, as described previously [95,111]. The *M. tuberculosis ΔrbpA attB*::tet-*rbpA*$_{Mtb}^{WT}$ strain was transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ (expressing RbpA$_{Mtb}^{WT}$ from a constitutive P*myc1-tetO* promoter; zeocin resistant) to replace the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of the *M. tuberculosis ΔrbpA* attB::tet-*rbpA*$_{Mtb}^{WT}$ strain. The transformants were selected with zeocin, and loss of the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct was confirmed by verifying their inability to grow in the presence of kanamycin. The *M. tuberculosis ΔrbpA*::tet-*rbpA*$_{Mtb}^{WT}$ strain transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ was named csm323. Csm323 was transformed with pMSG430-*rbpA*$_{Mtb}^{R79A}$, pMSG430-*rbpA*$_{Mtb}^{R88A}$, pMSG430-*rbpA*$_{Mtb}^{1–71}$, or pMSG430-*rbpA*$_{Mtb}^{72–111}$ (expressing RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$, RbpA$_{Mtb}^{1–71}$, or RbpA$_{Mtb}^{72–111}$, respectively, from a constitutive P*myc-tetO* promoter; kanamycin resistant) to replace the pDB19-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of csm323. The transformants were selected with kanamycin; when positive transformants in *M. tuberculosis* csm323 could not be obtained (as was the case for pMSG430-*rbpA*$_{Mtb}^{R79A}$,

pMSG430-*rbpA*$_{Mtb}^{R88A}$, pMSG430-*rbpA*$_{Mtb}^{1–71}$, and pMSG430-*rbpA*$_{Mtb}^{72–111}$ transformations), the mutations were deemed nonviable. (ii) *Mycobacterium smegmatis*. All *M. smegmatis* strains were derived from mc^2 155 and grown at 37°C in LB medium supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% Tween 80 (broth). The *M. smegmatis* merodiploid strain was constructed by integrating pMSG430-*rbpA*$_{Mtb}^{WT}$ into the *attB* site of mc^2 155. The *M. smegmatis* merodiploid strain was transformed with pDB88, with homology to mc^{2}155 nucleotides 3928650 to 3929246 and 3929589 to 3930405, to replace the endogenous *rbpA*, using two-step allelic exchange as described previously [112], thus generating Δ*rbpA*::tet-*rbpA*$_{Mtb}^{WT}$, which was named csm275. Csm275 was transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ to replace the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of the *M. smegmatis* Δ*rbpA attB*::tet-*rbpA*$_{Mtb}^{WT}$ strain. The transformants were selected with zeocin, and loss of the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct was confirmed by verifying their inability to grow in the presence of kanamycin. The *M. smegmatis* Δ*rbpA*::tet-*rbpA*$_{Mtb}^{WT}$ strain transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ was named csm291. Csm291 was transformed with pMSG430-*rbpA*$_{Mtb}^{WT}$, pMSG430-*rbpA*$_{Mtb}^{R79A}$, pMSG430-*rbpA*$_{Mtb}^{R88A}$, pMSG430-*rbpA*$_{Mtb}^{1–71}$, pMSG430-*rbpA*$_{Mtb}^{72–111}$, pMSG430-*rbpA*$_{Mtb}^{WT}$-FLAG, pMSG430-*rbpA*$_{Mtb}^{R79A}$-FLAG, pMSG430-*rbpA*$_{Mtb}^{R88A}$-FLAG, pMSG430-*rbpA*$_{Mtb}^{1–71}$-FLAG, or pMSG430-*rbpA*$_{Mtb}^{72–111}$-FLAG to replace the pDB19-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of csm291. Each FLAG tag repeated the sequence for FLAG twice (2X FLAG). The transformants were selected with kanamycin, and loss of the pDB19-*rbpA*$_{Mtb}^{WT}$ construct was confirmed by verifying their inability to grow in the presence of zeocin. When positive transformants in csm291 could not be obtained (as was the case for pMSG430-*rbpA*$_{Mtb}^{1–71}$ and pMSG430-*rbpA*$_{Mtb}^{1–71}$-FLAG transformations), the mutations were deemed nonviable. Csm291 strains transformed with pMSG430-*rbpA*$_{Mtb}^{R79A}$, pMSG430-*rbpA*$_{Mtb}^{R88A}$, pMSG430- rbpA$_{Mtb}^{72–111}$, pMSG430-*rbpA*$_{Mtb}^{WT}$-

FLAG, pMSG430-*rbpA*$_{Mtb}^{R79A}$-FLAG, pMSG430-*rbpA*$_{Mtb}^{R88A}$-FLAG, and pMSG430-*rbpA*$_{Mtb}^{72–111}$-FLAG were named csm322, csm314, csm328, csm313, csm329, csm327, and csm347, respectively.

Antibiotics and chemicals. In mycobacterial cultures, 20 µg/ml kanamycin and 12.5 µg/ml zeocin were used. In *E. coli cultures*, 40 µg/ml kanamycin, 50 µg/ml chloramphenicol, 50 µg/ml streptomycin, and 100 µg/ml ampicillin were used.

Western blotting and immunoprecipitation. For immunoprecipitation, 1-liter cultures were pelleted by centrifugation, resuspended in 20 ml of 1 phosphate-buffered saline (PBS) with complete protease inhibitor cocktail (Roche), and lysed with high-pressure (30 lb/in^2) cell disruption (CF model; Constant Systems, Daventry, UK). The lysate was treated with DNase I (New England BioLabs), added to anti-FLAG affinity gel (clone M2; Sigma-Aldrich, St. Louis, MO), and rotated overnight at 4°C. The protein-agarose matrix was washed three times with NP-40 buffer (10 mM sodium phosphate [pH 8.0], 150 mM NaCl, 1% Nonidet-40, 1 complete protease inhibitor cocktail). The immunoprecipitated protein complexes were eluted with 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 150 µg/ml FLAG peptide (SigmaAldrich), 1x complete protease inhibitor cocktail. Protein samples were mixed with SDS-PAGE loading buffer and run on a 4 to 12% Bis-Tris protein gel (Invitrogen). For the Western blot analysis, σ^A and σ^B were detected using a mouse monoclonal antibody against *E. coli* σ^{70} (clone 2G10; Neoclone, Madison, WI), RNAP β was detected using a mouse monoclonal antibody against *E. coli* RNAP β (clone 8RB13; Neoclone), and FLAG-tagged RbpA was detected using an anti-FLAG mouse monoclonal antibody (Sigma-Aldrich). Secondary LiCor IR Dye 800CW goat anti-mouse IgG polyclonal antibodies were used to detect the primary antibodies. Secondary antibody near-

infrared fluorescence was detected with the LiCore Odyssey version 3.0 imaging system, and band intensity was analyzed with Image Studio Lite version 4.0.

Protein purification for biochemical assays. Plasmids containing the *Mycobacterium tuberculosis* H37Rv genomic DNA encoding the different *M. tuberculosis* RNAP holoenzyme subunits were a gift from Jayanta Mukhopadhyay (Bose Institute, Kolkata, India) [113,114]. Expression was carried out in accordance with the method described by Banerjee et al. [114], with minor exceptions. Briefly, *E. coli* BL21(DE3) cells were transformed with plasmids pET-Duet-*rpoB-rpoC* (encoding the β and β' subunits), pAcYc-Duet-*sigA-rpoA* (encoding an N-terminal 10x His-tagged σ^A subunit and α subunit), and pCDF-*rpoZ* (encoding the ω subunit) and were grown in LB medium at 37°C to an optical density at 600 nm (OD_{600}) of 0.6 to 0.8. The culture was then treated with 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown overnight at 16°C. Cells were harvested via centrifugation (4,070 x g for 15 min at 4°C), and the resultant pellets were stored at 80°C. *M. tuberculosis* RNAP-σ^A holoenzyme was purified according to methods used previously for the *M. bovis* RNAP core complex [115]. *M. tuberculosis* RbpA constructs were cloned into pET-SUMO (Thermo Fisher Scientific) and transformed into *E. coli* BL21(DE3). Cultures were grown at 37°C to an OD_{600} of 0.8, and protein overexpression was induced with the addition of 0.5 mM IPTG overnight at 16°C. Cells were harvested by centrifugation (4,070 x g for 15 min at 4°C), and the cell pellets were stored at 80°C. The cells were resuspended in lysis buffer (50 mM NaH_2PO_4 [pH 8.0], 5 mM imidazole, 300 mM NaCl, 5 mM β-mercaptoethanol, protease inhibitor [Sigma-Aldrich]) and lysed by sonication at 4°C. Soluble lysate was separated from insoluble lysate by centrifugation (2,700 x g for 20 min at 4°C). RbpA was purified from the soluble lysate by Ni^{2+} affinity chromatography (Gold Biotechnology). Ni^{2+} columns were washed with wash buffer (50 mM Na_2HPO_4 [pH 8.0], 20

mM imidazole, 300 mM NaCl) until no protein was detected with NanoDrop spectrophotometer OD_{280} readings. RbpA was eluted from the Ni^{2+} affinity columns with elution buffer (50 mM Na_2HPO_4 [pH 8.0], 250 mM imidazole, 300 mM NaCl, 5 mM β-mercaptoethanol). The His-SUMO tag was cleaved from the RbpA constructs with His-Ulp1 protease during overnight dialysis at 4°C (20 mM Tris-HCl [pH 8.0], 250 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol). The His-SUMO tag and His-Ulp1 were separated from RbpA by a second round of Ni^{2+} affinity chromatography, and the cleaved RbpA was collected as the flowthrough fraction. Cleaved RbpA was dialyzed overnight at 4°C in storage buffer (20 mM Tris-HCl [pH 8.0], 250 mM NaCl, 1 mM β-mercaptoethanol), concentrated to approximately 200 µM (Vivaspin 20, molecular weight cutoff of 3,000; GE Healthcare), and stored at 80°C.

Preparation of fluorescent promoter DNA template. A Cy3-labeled promoter template of 150 bp 2+ nontemplate dT, containing nucleotides 1470151 to 1470300 of the *M. tuberculosis* Erdman genomic DNA, including the *rrn*AP3 promoter, was prepared as described previously [65,115].

Stopped-flow fluorescence assay. Stopped-flow experiments were performed as described previously [65,98,115], with notable exceptions. Prior to data acquisition, *M. tuberculosis* RNAP-σ^A holoenzyme, with or without RbpA protein, was incubated at 37°C for 10 min. All experiments were conducted with equal-volume mixing of 2 nM Cy3-labeled rrnAP3 promoter DNA with 70 or 200 nM *M. tuberculosis* RNAP-σ^A holoenzyme, with or without 4 µM RbpA protein. Thus, the final concentrations upon mixing were 1 nM DNA and 35 or 100 nM RNAP-σ^A holoenzyme, with or without 2 µM RbpA protein. Accounting for all contributions from protein storage buffers, the final reaction buffer conditions upon equal-volume mixing were as follows: 20 mM Tris (pH 8.0), 77.5 mM NaCl, 10 mM $MgCl_2$, 5 µM $ZnCl_2$, 20 M EDTA, 5% (vol/vol) glycerol,

1 mM dithiothreitol, and 0.1 mg/ml BSA. Experiments were performed with an SX-20 stopped-flow spectrophotometer (Applied Photophysics, Leatherhead, UK) with a dead time of 1 ms and a total shot volume of 100 µl. Samples were excited using a 535-nm fixed-wavelength light emitting diode (LED) light source with a 550-nm short pass filter, and emission was monitored using a 570-nm long pass filter. Data were collected at 37°C for 1,000 s by sampling 5,000 points over a logarithmic decay. Each protein condition is represented by the average of at least 5 shots obtained using multiple RNAP preparations, plotted as the fold change over DNA alone according to the formula $(F - F_0)/F_0$, where F_0 is the buffer-subtracted reading for DNA alone and F is the buffer-subtracted reading for DNA mixed with protein.

RNA-seq analysis. *M. smegmatis* strains csm275, csm322, csm314, and csm328 were cultured to an OD_{600} of 0.4 to 0.6, pelleted, resuspended in TRIzol (Thermo Fisher Scientific), and lysed by bead beating (FastPrep; MP Bio, Santa Ana, CA). RNA was extracted with chloroform, precipitated with isopropanol, and resuspended in water. RNA was treated with DNase I (Thermo Fisher Scientific), and RNA integrity and quality were analyzed with an Agilent bioanalyzer. rRNA was removed from samples using the Illumina Ribo-Zero rRNA removal kit. cDNA libraries were generated using an adapted Illumina TruSeq library preparation kit and were quality controlled by analysis of the cDNA size distribution with the Agilent TapeStation. cDNA libraries were pooled and sequenced in a single lane of an Illumina HiSeq 2000 Rapid Run flow cell with a 50-bp single-end read format. Sequencing reads were demultiplexed and converted to a FASTQ format using Illumina bcl2fastq script. Adapter sequences were trimmed from the raw reads, which were then aligned with the *M. smegmatis* mc^2 155 reference genome (GenBank accession number NC_008596) using the STAR aligner [116]. Sequence alignment map (SAM) files generated from alignments were converted to BAM files using SAMTools [117], and

aligned reads were counted per genome feature using the BioConductor package Subread featureCounts function [118]. Differential expression analysis and subsequent PCA were performed with BioConductor DESeq2 [119]. Venn diagrams were made with an online tool (https://www.stefanjol.nl/venny). Hypergeometric P values and enrichment values were calculated using an online calculator (http://systems.crump.ucla.edu/hypergeometric). The hypergeometric distribution describes the probability of k successes in s draws, without replacement, from a population of size N that contains exactly M successes. N was defined at the total number of differentially expressed genes in the two RbpA mutant constructs being compared, s was defined as the number of differentially upregulated or downregulated genes in one RbpA mutant included in the comparison, M was defined as the number of differentially upregulated or downregulated genes in the second RbpA mutant included in the comparison, and k was defined as the number of differentially upregulated or downregulated genes shared by the two RbpA mutants in the comparison.

qRT-PCR analysis. *M. smegmatis* strains csm275, csm322, csm314, and csm328 were cultured to an OD600 of 0.5 to 0.7, pelleted, resuspended in TRIzol (Thermo Fisher Scientific), and lysed by bead beating (FastPrep; MP Bio). RNA was extracted with chloroform, precipitated with isopropanol, and resuspended in water. MS2 bacteriophage RNA (Roche) was added to the bacterial RNA at a ratio of 1 ng of MS2 RNA per 1 billion bacteria, RNA was treated with DNase I (Thermo Fisher Scientific), and cDNA was synthesized with the Superscript III first-strand synthesis system (Invitrogen). qRT-PCR was performed with a SYBR green qPCR kit (Bio-Rad), and MSMEG_0281, MSMEG_5302, MSMEG_1215, MSMEG_3966, MSMEG_3297, MSMEG_3499, MSMEG_3855, MSMEG_1680, MSMEG_2259, MSMEG_4222, MSMEG_2758, MSMEG_2528, MSMEG_4497, MSMEG_6466,

MSMEG_6947, and MSMEG_2387 transcript levels were measured and normalized to spike-in MS2 RNA transcript levels. Primers are listed in Table S3 in the supplemental material.

Accession number(s). The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus [120] and are accessible through GEO Series accession number GSE107123.

Results

Individual RbpA structural domains are important for mycobacterial growth and viability.

To distinguish the roles of the RbpA structural domains in mycobacteria, we first engineered merodiploid strains of *M. tuberculosis* and *M. smegmatis* that expressed $rbpA_{Mtb}$ at the chromosomal *attB* site. The *M. smegmatis* and *M. tuberculosis* RbpA proteins are 92% identical. Expression of $rbpA_{Mtb}$ at the *attB* site allowed deletion of the endogenous *rbpA* gene in both *M. tuberculosis* and *M. smegmatis*, demonstrating that the RbpA protein from *M. tuberculosis* can substitute for the *M. smegmatis* RbpA protein to support viability. We then attempted to replace the $rbpA_{Mtb}$ gene at the *attB* site in *M. tuberculosis* and *M. smegmatis* with alleles encoding $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$, $RbpA_{Mtb}^{1–71}$, or $RbpA_{Mtb}^{72–111}$, using a previously described gene-swapping method [95,111]. The R79A mutation is within the BL and should disrupt DNA binding, the R88A mutation in the SID has been shown to weaken the affinity of RbpA for σ, the position 1 to 71 RbpA fragment is deleted for the BL and SID, and the position 72 to 111 RbpA fragment is deleted for the NTT and CD [22,109,110]. Using the gene-swapping approach, we found that none of the RbpA mutants could support viability in *M. tuberculosis*, demonstrating that *M. tuberculosis* is highly sensitive to any kind of disruption in RbpA function **(Figure 1a)**. In contrast, all of the mutant *rbpA* alleles except that encoding $RbpA_{Mtb}^{1–71}$ supported viability of *M. smegmatis*, thus providing us with a genetic system to study *M. tuberculosis* RbpA *in*

vivo by using *M. smegmatis* strains expressing RbpA$_{Mtb}^{WT}$, RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$, and RbpA$_{Mtb}^{72–111}$. The inability to obtain strains expressing the RbpA$_{Mtb}^{1–71}$ allele as the only *rbpA* allele demonstrated that the RbpA BL and SID are required for viability in mycobacteria.

To determine how each of these mutations in RbpA affected mycobacterial growth, the doubling times of *M. smegmatis* strains expressing the wild type (WT), RbpA$_{Mtb}^{WT}$, or mutant RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$, or RbpA$_{Mtb}^{72–111}$ were measured. The doubling times of the RbpA$_{Mtb}^{R79A}$ (4.3 h) and RbpA$_{Mtb}^{R88A}$ (4.4 h) strains were significantly longer than that of the RbpA$_{Mtb}^{WT}$ strain (3.2 h), indicating that the functions performed by the RbpA BL and SID are required for optimal *M. smegmatis* growth **(Figure 1a and 1b)**. Although the growth rate of the RbpA$_{Mtb}^{72–111}$ (3.9 h) strain trended lower than that of the RbpA*Mtb*^{WT} strain, this difference was not statistically significant, indicating that loss of the RbpA NTT and CD has only a mild effect on *M. smegmatis* growth.

To determine whether the mutations in RbpA affected the RbpA protein levels in *M. smegmatis*, we engineered *M. smegmatis* strains that expressed the C-terminally FLAG-tagged RbpA proteins RbpA$_{Mtb}^{WT}$-FLAG, RbpA$_{Mtb}^{R79A}$-FLAG, RbpA$_{Mtb}^{R88A}$-FLAG, and RbpA$_{Mtb}^{72–111}$-FLAG as the only copy of the *rbpA* product and we measured the levels of RbpA$_{Mtb}^{WT}$-FLAG, RbpA$_{Mtb}^{R79}$-FLAG, RbpA$_{Mtb}^{R88A}$-FLAG, and RbpA$_{Mtb}^{72–111}$-FLAG proteins in cell lysates by Western blot analysis. The levels of RbpA$_{Mtb}^{R88A}$-FLAG protein were significantly lower than the levels of RbpA$_{Mtb}^{WT}$-FLAG (data not shown). Therefore, the slower growth of the *M. smegmatis* strain expressing RbpA$_{Mtb}^{R88A}$ could in part be a result of lower levels of RbpA protein. The levels of RbpA$_{Mtb}^{72–111}$-FLAG protein were also significantly lower in cell lysates, compared to the levels of RbpA$_{Mtb}^{WT}$-FLAG. However, we found that this decrease in band

intensity was due to issues with the detection of RbpA$_{Mtb}^{72–111}$-FLAG with the anti-FLAG antibody. Therefore, it is unclear whether deletion of the RbpA NTT and CD decreases the levels of RbpA$_{Mtb}^{72–111}$-FLAG in cell lysates.

The RbpA SID is necessary and sufficient for association with RNAP.

Structural studies indicate that RbpA engages in four different macromolecular interactions in mycobacterial RNAP-promoter initiation complexes, i.e., (i) the RbpA NTT binding to RNAP β′ and σ, (ii) the RbpA CD binding to RNAP β′, (iii) the RbpA BL binding to DNA, and (iv) the RbpA SID binding to σ [22,106,109]; however, it is not known which of these interactions are required for the association of RbpA with the RNAP. To address this gap in knowledge, we performed coimmunoprecipitation experiments analyzing the amounts of σ^A, σ^B, and RNAP β subunit that coimmunoprecipitated with the RbpA-FLAG-tagged proteins. The levels of σ^A and σ^B coimmunoprecipitated with RbpA$_{Mtb}^{R88A}$-FLAG were dramatically reduced, compared to those coimmunoprecipitated with RbpA$_{Mtb}^{WT}$-FLAG, as expected based on the importance of R88 for σ binding [110] **(Figure 2a – 2c)**. In addition to the decreases in σ^A and σ^B levels, the levels of RNAP β coimmunoprecipitated with RbpA$_{Mtb}^{R88A}$-FLAG were significantly reduced **(Figure 2a and 2d)**. In crystallographic studies, the R88 in the RbpA SID is not positioned to bind directly to the core RNAP subunits; therefore, we conclude that the reduced β coimmunoprecipitated with RbpA$_{Mtb}^{R88A}$-FLAG is due to the reduced RbpA-σ interaction. This indicates that the interaction between the RbpA SID and the σ subunit is the primary determinant of the association of RbpA with the RNAP. In contrast, deletion of the NTT and CD (RbpA$_{Mtb}^{72–111}$) did not decrease the amounts of RNAP β, σ^A, or σ^B associated with RbpA **(Figure 2a – 2d)**. Therefore, despite the observations that the CD was positioned to interact with RNAP β′ and the NTT was positioned to interact with RNAP β′ and σ [67,106], these

interactions are not necessary for association with RNAP. Notably, although the levels of RNAP β, σ^A, and σ^B coimmunoprecipitated per molecule of $RbpA_{Mtb}^{72–111}$ appear to be increased in, we found that the differences were due to lower levels of $RbpA_{Mtb}^{72–111}$ detection by Western blot analysis (data not shown). $RbpA_{Mtb}^{R79A}$-FLAG coimmunoprecipitated similar levels of β and σ^A, compared to $RbpA_{Mtb}^{WT}$-FLAG **(Figure 2a, 2b and 2d)**. Coimmunoprecipitated levels of σ^B trended higher with $RbpA_{Mtb}^{R79A}$-FLAG but were not statistically significantly different from those observed with $RbpA_{Mtb}^{WT}$-FLAG **(Figure 2a and 2c)**. Collectively, our data show that the RbpA SID is both necessary and sufficient for interaction with RNAP.

RbpA mutants exhibit distinct kinetic phenotypes on the pathway to RP_o formation.

RbpA has been proposed to accelerate a forward kinetic step in the formation of RP_o, resulting in more stable RP_o at equilibrium [65,106]. A real-time fluorescence assay [65] was used to determine the effects of RbpA mutants on RP_o formation by the *M. tuberculosis* RNAP. Briefly, a Cy3 label was incorporated onto the +2 dT nucleotide, with respect to the +1 transcription start site, of the nontemplate strand of the *M. tuberculosis* rRNA *rrnA*P3 promoter [121]. The Cy3 label is positioned within the transcription bubble such that, upon opening of the promoter DNA, a 2-fold fluorescence enhancement is observed [81]; this allows quantitation of the kinetics of RP_o equilibration, by monitoring the change in fluorescence as a function of time, and the stability of RP_o, by using the equilibrium fluorescence value [65,98,115]. Incubating $RbpA_{Mtb}^{WT}$ at a saturating concentration (2 μM) with 35 nM *M. tuberculosis* RNAP-σ^A holoenzyme and the Cy3-labeled *rrnA*P3 promoter resulted in a greater amount of RP_o at equilibrium than observed with RNAP-σ^A holoenzyme and the *rrnA*P3 promoter alone, consistent with the known role of RbpA in stabilizing the otherwise unstable mycobacterial RNAP open complex [10,65]. When the same concentrations of $RbpA_{Mtb}^{R79A}$ and $RbpA_{Mtb}^{R88A}$ were added, no enhancement of the amount of

RP_o at equilibrium over RNAP-σ^A holoenzyme and the *rrnA*P3 promoter alone was observed **(Figure 3a and 3b)**, demonstrating the importance of these residues. For a qualitative description of the kinetics, we calculated $t_{1/2}$ values (the time required to reach the midpoint of the final equilibrium fluorescence). Interestingly, these mutants exhibited approximately 3-fold faster kinetics ($RbpA_{Mtb}^{R79A}$ $t_{1/2}$ of 14.8 ± 1.1 s and $RbpA_{Mtb}^{R88A}$ $t_{1/2}$ of 16.4 ± 1.1 s), compared with the RNAP-σ^A holoenzyme and the *rrnA*P3 promoter alone ($t_{1/2}$ of 43 ± 2 s) **(Figure 3a and 3c)**. The finding of faster kinetics accompanied by no change in the equilibrium fluorescence value suggests that these mutants retain the ability to stabilize the transition state on the pathway to RP_o but have lost the ability to stabilize RP_o itself. This behavior is analogous to the classic model for enzyme activity [122], in which the transformation of substrate to product is accelerated without changes in the final equilibrium between the two states. In this scenario, the mutant RbpA proteins may increase the rate of opening and the rate of closing equally, such that the ratio of rates remains constant. These results suggest that the interactions between RbpA and both the promoter DNA (R79) and σ factor (R88) are essential for RP_o stabilization and that RbpA is still capable of catalyzing promoter opening even in the presence of these mutations.

To further investigate the domain requirements for RP_o stabilization, we repeated the experiments described above with $RbpA_{Mtb}^{1–71}$ (containing the NTT and CD) and $RbpA_{Mtb}^{26–71}$ (containing the CD only), and we observed minimal enhancement in RP_o stability and an identical rate of RP_o equilibration ($t_{1/2}$ of 42 ± 3 s), relative to the RNAP-σ^A holoenzyme and the *rrnA*P3 promoter alone, indicating that the NTT and CD are unable to affect RP_o stability on their own **(Figure 3a – 3c)**. Conversely, $RbpA_{Mtb}^{72–111}$ showed the greatest amount of RP_o at equilibrium, even higher than that of $RbpA_{Mtb}^{WT}$, with kinetics ($t_{1/2}$ of 7.8 ± 0.9 s) similar to those of $RbpAMtb^{WT}$ ($t_{1/2}$ of 6.9 ± 0.5 s). The finding that $RbpAMtb^{72–111}$ exhibits similar

kinetics but a greater amount of RP_o at equilibrium, compared with $RbpA_{Mtb}^{WT}$, raises the possibility that the NTT and CD negatively affect RbpA activity under these conditions. To determine whether it was the NTT and/or the CD that antagonized RbpA-mediated RP_o stabilization, we assayed an RbpA protein with deletion of just the NTT ($RbpA_{Mtb}^{26–111}$). $RbpA_{Mtb}^{26–111}$ yielded a greater fold change in fluorescence than did $RbpA_{Mtb}^{WT}$ but smaller change than did $RbpA_{Mtb}^{72–111}$, suggesting that both the CD and NTT are responsible for the antagonistic effect on RbpA-dependent RP_o stability. $RbpA_{Mtb}^{26–111}$ exhibited approximately 2-fold slower kinetics of RP_o equilibration ($t_{1/2}$ of 17.6 ± 1.2 s) than did $RbpA_{Mtb}^{72–111}$ ($t_{1/2}$ of 7.8 ± 0.9 s) and $RbpA_{Mtb}^{WT}$ ($t_{1/2}$ of 6.9 ± 0.5 s). One possibility consistent with this observation is that, in the presence of the rest of the domains, the NTT decreases the amount of RP_o at equilibrium by increasing a reverse rate leading toward the RNAP-promoter closed complex (RP_c). Importantly, performing these experiments with multiple RNAP concentrations suggests that the effect of each RbpA construct is limited by DNA-bound kinetic intermediates and not the rates of association and dissociation of RNAP to and from promoter DNA. Taken together, these results suggest that residues R79 and R88 are essential for RP_o stabilization and that the NTT and CD can inhibit RP_o formation.

Truncation of the RbpA NTT/CD and mutations in the RbpA BL and SID result in distinct gene expression changes in *M. smegmatis*.

To determine how the individual RbpA domains contribute to gene expression, we performed RNA-sequencing (RNA-seq) experiments with cultures of *M. smegmatis* expressing $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$, $RbpA_{Mtb}^{72–111}$, or $RbpA_{Mtb}^{WT}$. The only previous analysis of this type focused on the gene expression profiles that resulted from deletion of the RbpA NTT and CD in *M. smegmatis*, but it did not investigate the roles of the other RbpA domains [106]. Principal-

component analysis (PCA) of the RNA-seq data was performed and provided a general overview of how gene expression patterns among the RbpA mutants clustered in relationship to each other. Three distinct sample clusters were apparent from the PCA results (**Figure 4a**), indicating three different gene expression patterns. The first cluster included the three RbpA$_{Mtb}^{WT}$ replicates, the second cluster included the three RbpA$_{Mtb}^{72–111}$ replicates, and the third cluster included the replicates from both RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$. The PCA results indicate that loss of the RbpA NTT/CD affects a gene subset that is different from the genes affected by mutations in the RbpA BL and SID. The number of genes significantly (adjusted *p* values of <0.05) upregulated or downregulated 2-fold in the RbpA mutants varied, with 766 genes being differentially expressed in RbpA$_{Mtb}^{72–111}$, compared to 199 genes in RbpA$_{Mtb}^{R79A}$ and 244 genes in RbpA$_{Mtb}^{R88A}$ (**Figure 4b**).

Consistent with the PCA results, there was significant overlap in upregulated and downregulated genes between the RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ strains (**Figure 4c and 4d**), indicating that the SID and BL perform functions that contribute to the expression of a common subset of *M. smegmatis* genes. Also consistent with the PCA results, the upregulated and downregulated genes in RbpA*Mtb*$^{72–111}$ had little overlap with those in either RbpA$_{Mtb}^{R79A}$ or RbpA$_{Mtb}^{R88A}$ (**Figure 4c and 4d**). Therefore, the number of shared downregulated or upregulated genes between RbpA$_{Mtb}^{72–111}$ and either RbpA$_{Mtb}^{R79A}$ or RbpA$_{Mtb}^{R88A}$ was under enriched (**Figure 4c and 4d**).

Given that RbpA stabilizes RNAP-σ^A-*rrnA*P3 RP_o *in vitro* and the R79 and R88 residues are essential for this activity, it might be expected that the RbpA BL and SID cooperate to activate transcription from all promoters that RbpA regulates. Similarly, the ability of the NTT and CD to antagonize RbpA-mediated RP_o stabilization would lead to the hypothesis that

expression from RbpA-regulated genes would be increased in their absence. However, this was not supported by the RNA-seq data, in which similar numbers of transcripts were upregulated and downregulated in each RbpA mutant (**Figure 4b**). These data could mean that domains within RbpA can promote both activation and repression of gene expression. However, it is also possible that there was general downregulation or upregulation of gene expression in the RbpA mutants that we were unable to detect due to the addition of equal amounts of RNA from each strain into the sequencing reaction. To explore this possibility, we performed spike-in experiments [123] in which we isolated RNA from cultures of *M. smegmatis* expressing RbpA*Mtb*^{R79A}, RbpA*Mtb*^{R88A}, RbpA*Mtb*$^{72–111}$, or RbpA*Mtb*^{WT} and added 1 ng of MS2 bacteriophage RNA (Roche) per 1 billion bacterial cells to the RNA samples. cDNA was generated for each sample, and quantitative reverse transcription-PCR (qRT-PCR) was performed to determine transcript levels for 16 *M. smegmatis* genes relative to MS2 RNA, which was used as a proxy to represent cell number. The 16 *M. smegmatis* genes analyzed included genes that were significantly upregulated or downregulated in RbpA mutants during RNA-seq experiments. When results were normalized to MS2 RNA levels, all 16 genes, including the genes considered highly upregulated in the RNA-seq analysis, were downregulated in RbpA$_{Mtb}^{72–111}$ compared to RbpA$_{Mtb}^{WT}$, suggesting that overall transcript levels in RbpA$_{Mtb}^{72–111}$ are decreased (**Figure 4e**). Therefore, despite the findings that deletion of the NTT and CD had only a mild effect on the growth rate (**Figure 1b and 1c**) and enhanced RP_o stabilization activity *in vitro* (**Figure 3a and 3b**), the NTT and CD are required for WT levels of gene expression in *M. smegmatis*. In contrast, qRT-PCR results for the RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ mutants were similar to the RNA-seq results, indicating that RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ mutants do indeed lead to both upregulation and downregulation of gene

expression. When we analyzed the genes that were most upregulated or downregulated with the $RbpA_{Mtb}^{R79A}$ and $RbpA_{Mtb}^{R88A}$ mutants, they fell into multiple diverse functional classes, indicating that RbpA activity likely affects multiple cellular processes.

Discussion

In this study, we investigated the functions of the individual RbpA structural domains to gain insight into the complex *in vivo* roles of RbpA. To study the roles of the RbpA NTT and CD, we truncated the N-terminal 71 amino acids of RbpA. The role of the RbpA BL was probed using a point mutation at R79, which has been implicated in the interaction between RbpA and DNA [22]. Finally, we investigated the RbpA SID by using a point mutation at R88, which is one of the key residues needed for the interaction between RbpA and [110] but had yet to be studied in mycobacteria in *vivo*. We found that the function of each RbpA structural domain is required for *M. tuberculosis* viability and wild-type growth rates in *M. smegmatis* and disruption of the RbpA BL and SID functions causes a more severe growth defect than loss of the NTT and CD (**Figure 1**) Our data indicate that M. tuberculosis has a more stringent requirement for RbpA activity, similar to what we observed for CarD [94,95].

We determined that the RbpA SID interaction with is the only interaction required for the association of RbpA with the RNAP; the RbpA R88A substitution resulted in not only loss of the interactions with σ^A and σ^B but also almost complete loss of association with the core RNAP subunit (**Figure 2**). In contrast, deletion of the NTT and CD did not negatively affect the association of RbpA with RNAP, suggesting that the RbpA NTT and CD serve functions distinct from interaction with RNAP. The RbpA R88A substitution also resulted in decreased RbpA protein levels. Previous studies investigating CarD mutants with altered affinities for the RNAP found that CarD protein levels correlated with CarD affinity for the RNAP [98]. Our data showing

that RbpA$_{Mtb}^{R88A}$ has a lower affinity for the RNAP and is present in lower abundance in the cell supports a model in which RbpA protein levels are also affected by its ability to interact with the RNAP. CarD was shown to be a target of the Clp protease in M. tuberculosis and, similarly, RbpA levels were >2-fold higher in a M. tuberculosis strain lacking Clp protease subunits, suggesting that RbpA protein levels may also be regulated by the Clp protease [124].

Previous studies investigating the effect of RbpA on RP_o stability reported that R79 is required for RP_o stabilization, whereas both the NTT and CD are dispensable [65,106]. We have expanded on these findings by determining that the RbpA- interaction is required for enhanced RP_o stability and the NTT and CD antagonize this activity (**Figure 3**). Furthermore, our results suggest that the effect of RbpA on the kinetics of RP_o equilibration can be differentiated from its effect on equilibrium levels of RP_o and that RbpA can affect both forward and reverse rates on the pathway to RP_o, at the concentrations tested. The similar effects of RbpA BL and SID mutations on RP_o stabilization (**Figure 3**) mirror the significant overlap in the expression profiles of RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ (**Figure 4**). In contrast, the truncation of NTT and CD, which affects RP_o stability differently than mutations in RbpA BL and CD, results in an expression profile significantly different from that of RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$. We have found that RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ mutants can result in both upregulation and downregulation of transcript levels in *M. smegmatis*, depending on the gene. Upregulation of gene expression in RbpA mutants could be due to direct effects with RbpA acting as a repressor in some promoter contexts, due to differences in basal initiation kinetics. However, this observation could also be explained by indirect effects with RbpA enhancing the expression of a transcription factor that represses the expression of a set of genes. Future studies that expand analysis of RbpA past the limited promoters that have been explored in vitro will be necessary to address these

possibilities. Our data from spike-in control qRT-PCR experiments suggest that gene expression is globally downregulated in the *M. smegmatis* $RbpA_{Mtb}^{72–111}$ mutant. This suggests that the NTT and CD are required for efficient gene expression, and it complicates interpretations of the $RbpA_{Mtb}^{72–111}$ RNA-seq data in this study. This finding may also have an impact on a previously published RNA-seq data set for the *M. smegmatis* $RbpA_{Mtb}^{72–111}$ strain [106]. When we compared our RNA-seq data set for $RbpA_{Mtb}^{72–111}$ with the previously reported data, we found that there was no significant overlap in genes that registered as upregulated or downregulated. This could be due to a difference in the culturing methods used in the two studies and/or it could be related to the finding that gene expression in general is less robust. How the NTT and CD mechanistically promote efficient gene expression while antagonizing RP_o stability on the rrnAP3 promoter in vitro remains an open question for future studies.

Acknowledgements

C.L.S. and E.A.G. were supported by grant GM107544 from the NIH. C.L.S. was also supported by a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease Award. J.P. and A.L.G. were supported by NIGMS grant GM007067. A.L.G. was also supported by a Stephen I. Morse Graduate Fellowship. G.S.-C. was supported by NIH grant R25HG006687. D.J. was supported by a Gary K. Ackers Fellowship and an Elliot L. Elson Education and Training Fellowship. J.J.M. was supported by NIH training grant AI007172-36A1. Purchase of the stopped-flow fluorescence equipment was made possible by equipment supplement 3R01GM107544-04S1 from the NIH. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis; the center is partially supported by NCI Cancer Center Support grant P30CA91842 to Siteman Cancer Center and by ICTS/CTSA grant UL1TR000448 from the National Center for

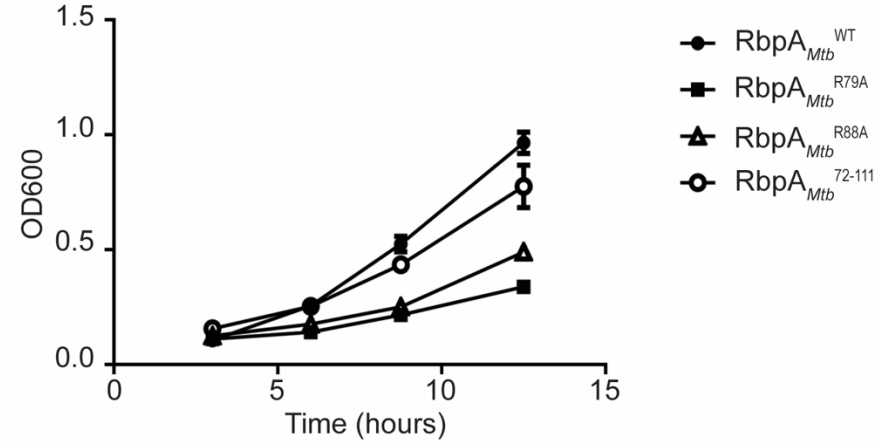
Research Resources, a component of the NIH, and the NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the NIH.

Figures

a

Mtb RbpA Protein	Viable *M. smegmatis*	Viable *M. tuberculosis*	Predicted Function Affected
$RbpA_{Mtb}^{R79A}$	Yes	No	DNA binding
$RbpA_{Mtb}^{R88A}$	Yes	No	σ binding
$RbpA_{Mtb}^{72-111}$	Yes	No	RNAP β' binding
$RbpA_{Mtb}^{1-71}$	No	No	DNA and σ binding

b



c

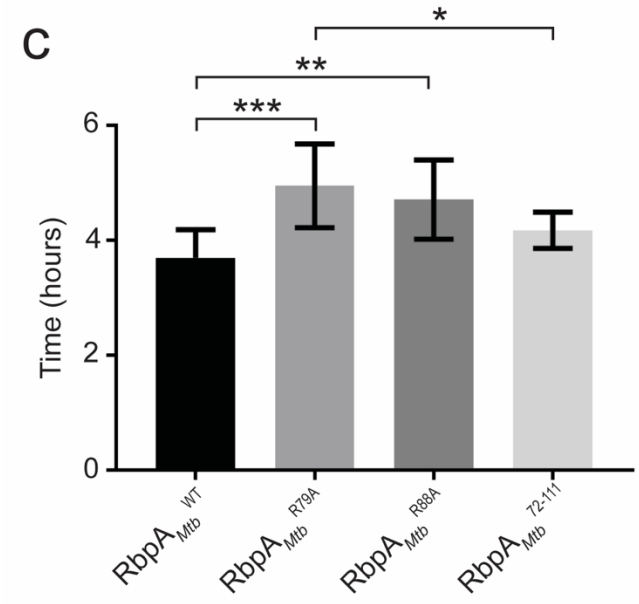


Figure 1. Individual RbpA structural domains are important for mycobacterial growth and viability

a. Viability of mutant RbpA constructs, $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ and $RbpA_{Mtb}^{72\text{-}111}$ in *M. tuberculosis* and *M. smegmatis*.

b. Representative growth of *M. smegmatis* strains expressing $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$.

c. Doubling times of *M. smegmatis* strains expressing $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$ calculated from growth curves as shown in (b). Results are plotted as ± standard deviation. Statistical significance was analyzed by analysis of variance (ANOVA) and Tukey's multiple comparison test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.01$.

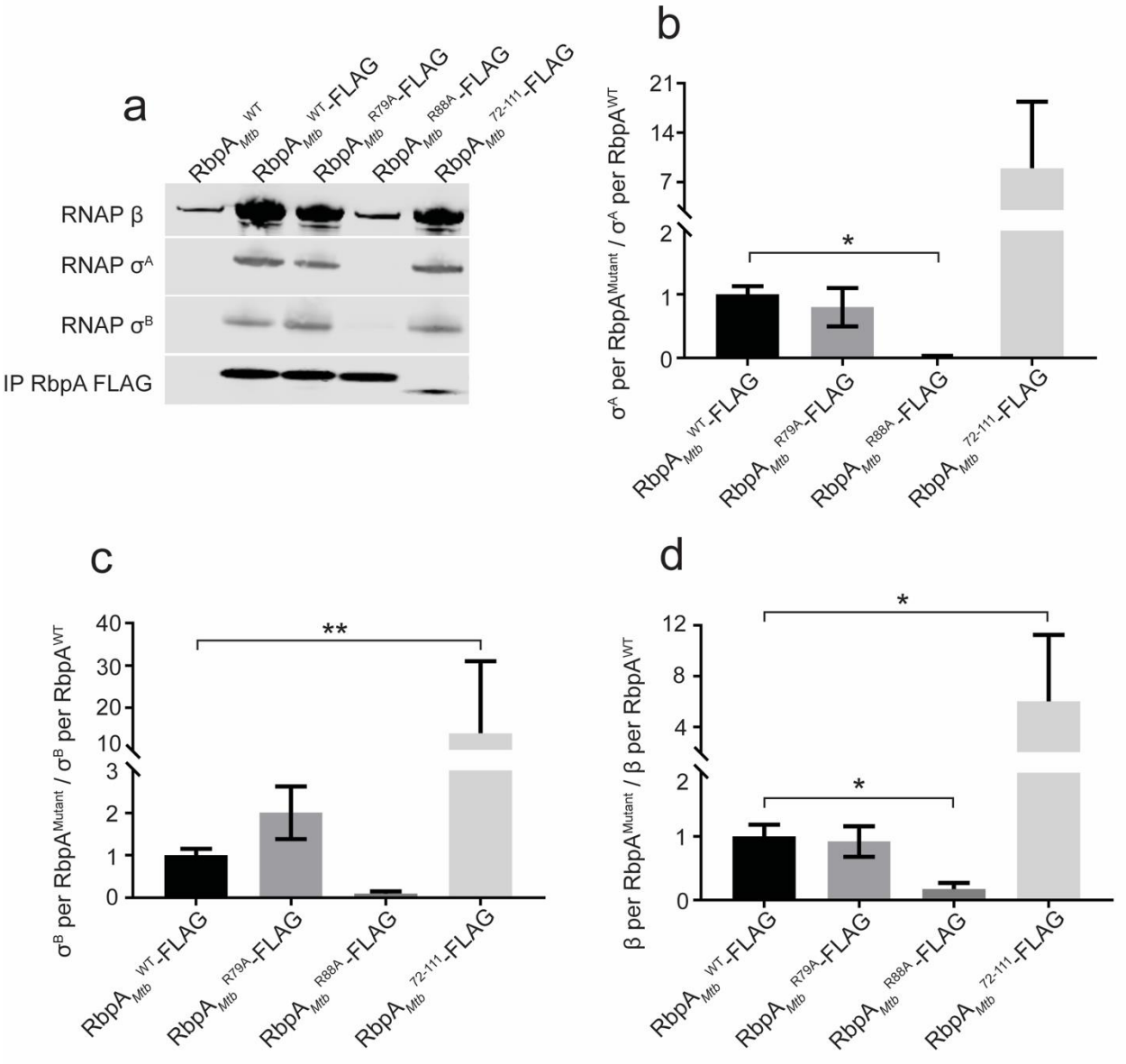


Figure 2. The RbpA SID is necessary and sufficient for association with RNAP

a. Western blot analysis of lysates immunoprecipitated for FLAG-tagged RbpA. Monocolonal antibodies specific for FLAG were used to detect $RbpA_{Mtb}$ – FLAG protein variants (bottom row). RNAP β, and both σ^A and σ^B were detected using a monoclonal antibody specific for a shared epitope in *E. coli* σ^{70}.

b – d. Amounts of σ^A (b), σ^B (c) and RNAP β (d) coimmunoprecipitated by RbpA, based on band intensity, and expressed as the ratio of σ^A, σ^B or RNAP β to RbpA, with eight replicates for each strain. Results are shown as means ± standard deviations. Statistical significance was determined by one-way ANOVA and Kruskal-Wallis multiple-comparison test. *, $p < 0.05$; **, $p < 0.01$.

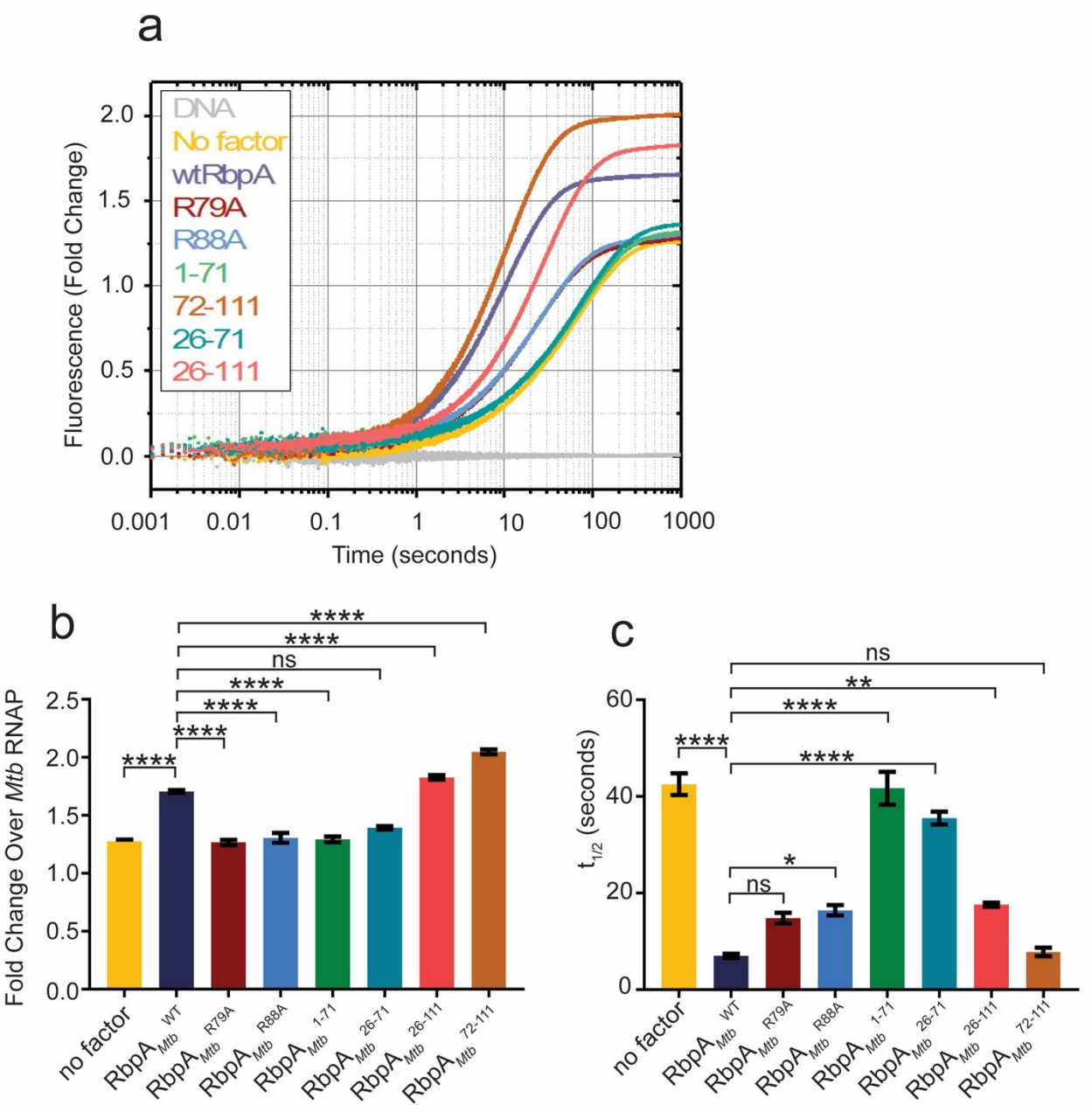


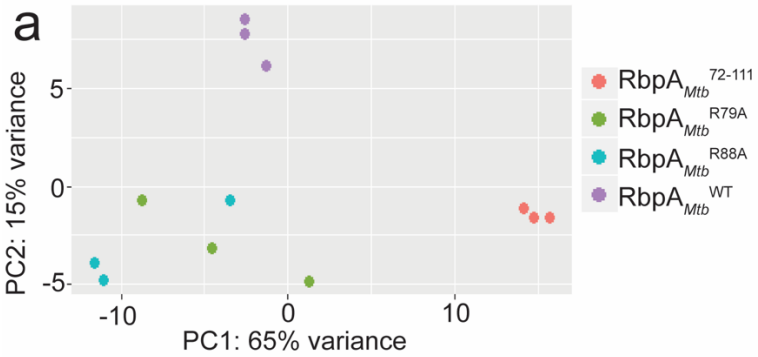
Figure 3. RbpA mutants exhibit distinct effects on RP_o formation.

a. Fluorescence fold changes, compared to DNA alone, which were was used to monitor RP_o formation and stability in real time, using fixed amounts of *M. tuberculosis* RNAP - σ^A (35 nM), Cy3-labeled (+ 2 thymine nontemplate strand) *M. tuberculosis rrn*AP3 promoter DNA (1 nM), and RbpA (2 µM). Time courses are shown as an average of at least 5 replicates.

b. Total fluorescent fold changes, normalized to RNAP- σ^A on *rrn*AP3 alone, for all RbpA constructs.

c. t1/2 values, calculated as the time required to reach one-half of the final fluorescence intensity, for each sample. For panels (b) and (c), means standard errors of the means are plotted. Statistical significance was analyzed by ANOVA and Tukey's multiple-comparison test. *, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$; ns, not significant. Only comparisons between RbpAWT and each of the RbpA mutant constructs are shown in the figure

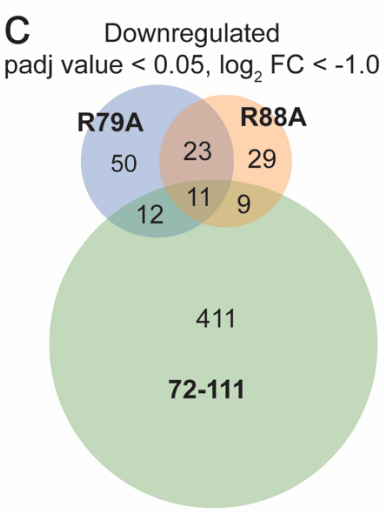
a



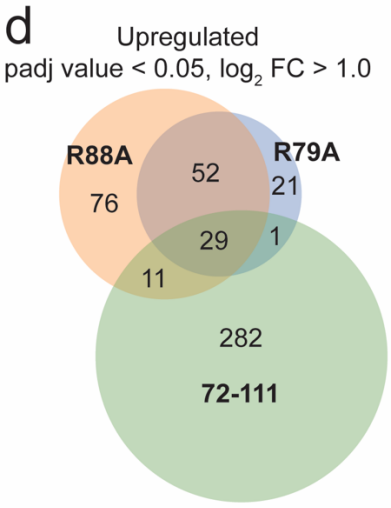
b

Direction of dysregulation	$RbpA_{Mtb}^{R79A}$	$RbpA_{Mtb}^{R88A}$	$RbpA_{Mtb}^{72-111}$
upregulated	103	168	322
downregulated	96	72	443

c



d



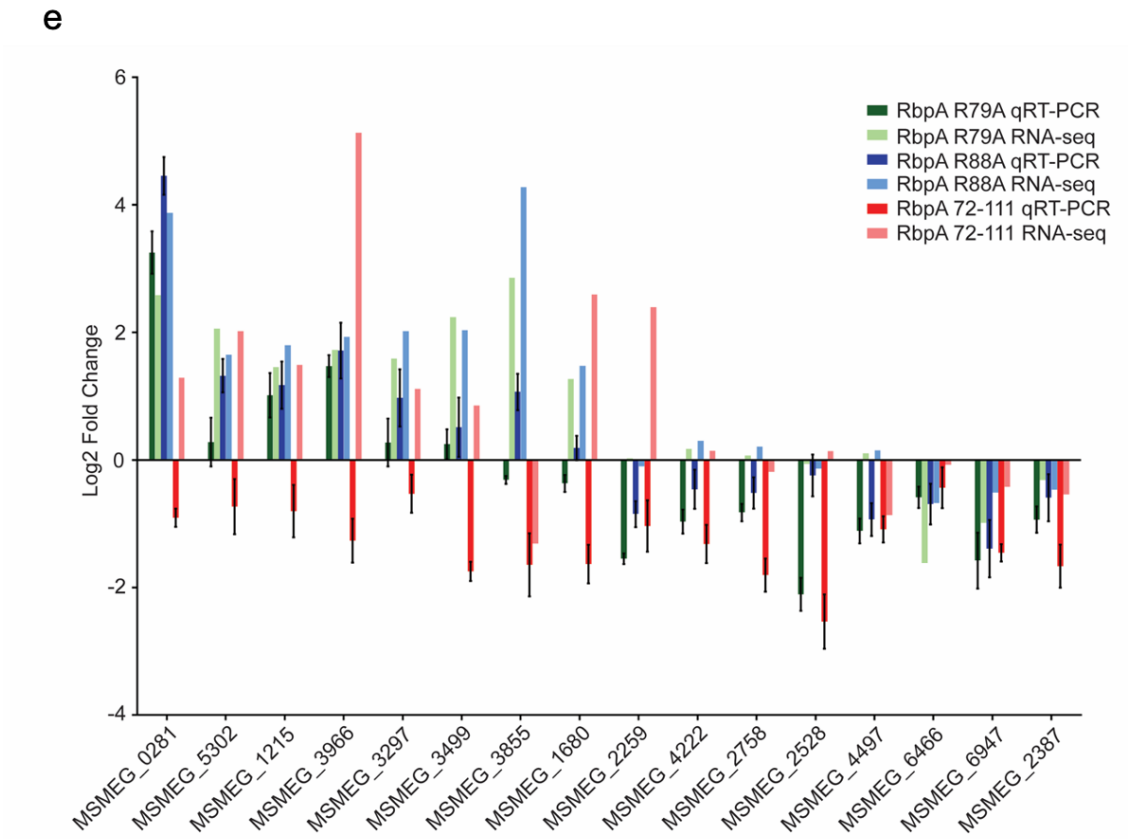


Figure 4. Truncations of RbpA NTT/CD and mutations in the RbpA BL and SID result in distinct gene expression changes in *M. smegmatis*

a. PCA results showing samples distances across two principal components, generated using read counts of RNA collected from *M. smegmatis* expressing RbpA$_{Mtb}^{WT}$, RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$ or RbpA$_{Mtb}^{72\text{-}111}$, mapped to the *M. smegmatis* mc^{2}155 genome and normalized with regularized logarithmic transformation. Each represents one of three replicates for RbpA$_{Mtb}^{WT}$, RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$ or RbpA$_{Mtb}^{72\text{-}111}$.

b. Number of genes significantly (adjusted p values of < 0.05) upregulated or downregulated 2-fold in *M. smegmatis* expressing RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$ or RbpA$_{Mtb}^{72\text{-}111}$, relative to *M. smegmatis* expressing RbpA$_{Mtb}^{WT}$. FC, fold change.

c. Venn diagram showing overlap of the genes downregulated 2-fold (adjusted p value of < 0.05) in *M. smegmatis* expressing expressing $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$, relative to *M. smegmatis* expressing $RbpA_{Mtb}^{WT}$.

d. Venn diagram showing overlap of the genes upregulated 2-fold (adjusted p value of < 0.05) in *M. smegmatis* expressing expressing $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$, relative to *M. smegmatis* expressing $RbpA_{Mtb}^{WT}$.

e. qRT-PCR and RNA-seq $\log_2$ fold changes for 16 genes in *M. smegmatis* expressing $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$, relative to *M. smegmatis* expressing $RbpA_{Mtb}^{WT}$. Transcript levels were normalized to a MS2 RNA spike-in control that was added at a constant level of 1ng/billion cells. Means ± standard errors of the means of three replicates are shown for each *M. smegmatis* strain.

Chapter 3: Molecular Dissection of RbpA-Mediated Regulation of Transcription Initiation and Fidaxomicin Sensitivity in Mycobacteria

Jerome Prusa, Dennis X. Zhu, Drake Jensen, Ana Ruiz Manzano, Eric A. Galburt, Christina L. Stallings

A version of this chapter is currently under revision at the Journal of Biological Chemistry

Abstract

RNA polymerase (RNAP) binding protein A (RbpA) is essential for mycobacterial viability and regulates transcription initiation by increasing RNAP-promoter open complex (RP_o) stability. RbpA consists of four structural domains: a N-terminal tail (NTT), core domain (CD), basic linker (BL), and sigma interaction domain (SID). The roles of the BL and SID have been studied extensively, whereas, understanding of the NTT and CD is still limited. Truncation of the RbpA NTT and CD increases RP_o stabilization by RbpA *in vitro*, implying that these domains antagonize this activity of RbpA. Structural studies show that the NTT and CD are positioned near multiple RNAP-σ^A holoenzyme functional domains, suggesting that the RbpA NTT and CD could have a number of effects on RNAP activity and it is unclear what contacts mediate the antagonism of RP_o stability. In addition, structural studies predict that the RbpA NTT contributes contacts to the binding site for the antibiotic fidaxomicin (Fdx) on RNAP. Deletion of the NTT results in a decrease in *Mycobacterium smegmatis* sensitivity to Fdx, but whether this is caused by a loss of contacts with Fdx has yet to be tested. To address these gaps in knowledge, we generated a panel of *rbpA* mutants with single amino acid substitutions in conserved residues within the NTT to probe what residues are involved in regulating Fdx activity and RP_o stability. We find that the NTT along with other RbpA domains and CarD contribute to Fdx activity *in vivo*, while *in vitro* the RbpA NTT residues that directly interact with Fdx are only partially responsible for RbpA NTT dependent Fdx activity. We identify roles for conserved RbpA NTT residues in RbpA NTT antagonism of RP_o stabilization and determine that the loss of RbpA NTT decreases full-length transcription. This finding provides the first *in vitro* evidence that RbpA RP_o stability can directly inhibit full-length transcription supporting the model that RbpA activities both activate and inhibit full-length transcription.

Introduction

Compared to RbpA BL and SID much less is known about the functions performed by the RbpA NTT and CD. Deletion of the RbpA NTT increases RbpA RP_o stabilizing activity and deletion of both the RbpA NTT and CD further increases RbpA RP_o stabilizing activity, indicating that both domains antagonize RbpA RP_o stabilizing activity [125]. Structural analysis of RbpA bound to the *M. tuberculosis* RNAP-σ^A RP_o shows that the RbpA NTT is positioned near the RNA exit channel, possibly contacting the RNAP β switch 3 region (Sw3), β flap, β' lid, σ^A region 3.2 ($\sigma^A_{3.2}$, also referred to as the σ "finger" domain), and the β' zinc binding domain (ZBD), while the RbpA CD is positioned near the RNAP β' zipper and RNAP β' ZBD [67,106]. These RNAP structural domains have been characterized to varying levels in *E. coli*, which lacks RbpA. The RNAP β Sw3 is one of five switch regions that are thought to undergo conformational changes during transcription initiation [126]. RNAP β Sw3 is positioned near the template DNA -3 and -4 nucleotides, raising the possibility that RNAP β Sw3 could play a role in DNA template strand positioning [6]. The RNAP β flap, which includes the flap tip helix that interacts with σ region 4, is important for positioning σ region 4 for interaction with the -35 element of the promoter [127] and represents a common binding interface for transcription factors that directly interact with σ [128,129]. The RNAP β' lid separates the RNA/DNA hybrid as part of the RNA exit channel and is required for RP_o stability and transcription in *E. coli* and *Thermus aquaticus* [130,131]. RNAP $\sigma^{70}_{3.2}$ plays a role in initiating nucleotide triphosphate (iNTP) binding by positioning the DNA template strand for interaction with -4 and -5 nucleotides of the DNA template strand, which affects abortive transcription and promoter escape [28–31]. Both the RNAP β' ZBD and β' zipper facilitate RP_o formation on promoters with -35 elements that form weak

interactions with σ by making promoter contacts within the spacer region between the -10 and -35 motifs [132,133].

The positioning of the RbpA NTT and CD near multiple different structural and functional domains of the RNAP-σ^A holoenzyme implies that the RbpA NTT and CD could impact RNAP activity through a number of mechanisms. However, it is unclear what contacts between the RbpA NTT/CD and the RNAP mediate the antagonism of RP_o stability. In addition, structural studies indicate that the RbpA NTT is positioned in the RNAP-σ^A holoenzyme complex in such a way that it contributes to the binding site for the antibiotic fidaxomicin (Fdx) [67]. Fdx inhibits transcription initiation by binding the RNAP and blocking the closing of the RNAP clamp that occurs during RP_o formation [67,88]. Deletion of the RbpA NTT decreases sensitivity to Fdx *in vitro* and *in vivo* [67], which is proposed to be due to the loss of RbpA's contribution to the RNAP-Fdx binding interface. However, given that RbpA NTT also decreases RP_o stability [106,125] and is predicted to interact with the $\sigma^A_{3.2}$, which is known to affect Fdx activity [67,134], it is possible that RbpA may impact Fdx activity by additional mechanisms. In this study, we interrogate the roles played by conserved residues within the NTT in RbpA-dependent Fdx sensitivity and RP_o stabilization. In addition, we link NTT antagonism of RP_o stability to increased full length transcript production, suggesting that decreasing RP_o stability could allow for the transition of the RNAP-RbpA complex through transcription initiation and promoter escape.

Experimental Procedures

Media and bacterial strains. All *Mycobacterium smegmatis* strains were derived from mc^2155 and grown at 37°C in LB medium supplemented with 0.5% dextrose, 0.5% glycerol and 0.05% Tween 80. *M. smegmatis* strains expressing $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7A}$,

RbpA$_{Mtb}^{R7E}$, RbpA$_{Mtb}^{R10A}$, RbpA$_{Mtb}^{S15A}$, RbpA$_{Mtb}^{E17A}$ and RbpA$_{Mtb}^{R10A/E17A}$, RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$, RbpA$_{Mtb}^{26-111}$, RbpA$_{Mtb}^{72-111}$, RbpA$_{Msm}^{28-114}$ and RbpA$_{Msm}^{72-114}$ were engineered using pMSG430 plasmids that express each *rbpA* allele from a constitutive P*myc1-tetO* promoter and integrate into the the *attB* site of the *M. smegmatis* Δ*rbpA attB::tet-rbpA* strain previously described [95,111,125]. The *M. smegmatis* Δ*rbpA attB::tet-rbpA* strains expressing RbpA$_{Mtb}^{R4A}$, RbpA$_{Mtb}^{R4E}$, RbpA$_{Mtb}^{L6A}$, RbpA$_{Mtb}^{R7A}$, RbpA$_{Mtb}^{R7E}$, RbpA$_{Mtb}^{R10A}$, RbpA$_{Mtb}^{S15A}$, RbpA$_{Mtb}^{E17A}$, RbpA$_{Mtb}^{R10A/E17A}$, RbpA$_{Msm}^{28-114}$, RbpA$_{Msm}^{72-114}$, RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ were named csm455, csm461, csm456, csm457, csm458, csm451, csm462, csm450 and csm498, csm510, csm511, csm322 and csm314 respectively.

Protein Preparation for biochemical assays. Plasmids containing the *M. tuberculosis* H37Rv genomic DNA encoding the different *M. tuberculosis* RNAP holoenzyme subunits were a gift from Jayanta Mukhopadhyay (Bose Institute, Kolkata, India) [113,114]. Expression and purification were carried out in accordance with the methods described previously in [125]. Recombinant *M. tuberculosis* RbpA proteins were purified from *E. coli* as previously described [125]. RbpA was stored at -80^{o}C 150 mM NaCl, 20 mM Tris pH 8.0, and 1 mM β-mercaptoethanol. *M. tuberculosis* RNAP-σ^{A} holoenzyme was stored at -80^{o}C in 50% glycerol, 10 mM Tris pH 7.9, 200 mM NaCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM $MgCl_2$, and 20 μM $ZnCl_2$.

Fidaxomicin Zone of Inhibition. *M. smegmatis* cultures were grown to OD_{600} = 0.4 - 0.8. Based on the approximation that OD_{600} = 1.0 is equivalent to 5 x 10^8 mycobacteria, 2.5 x 10^8 cells were collected, resuspended in 100 μl of LB, and plated on LB agar plates. Whatman filter paper disks were applied to the plates and 10 μl of 100 μM, 250 μM, or 500 μM fidaxomicin (Selleck Chemicals) resuspended in DMSO or DMSO alone were added to the Whatman filter paper

disks. The plates were incubated at 37ºC for 48 hours and the zones of inhibition were measured. The zone of inhibition for each replicate at each drug concentration is the average of four measurements approximately 90º apart.

3-nucleotide *in vitro* transcription assay. Linear 150bp dsDNA template containing the *M. tuberculosis rrn*AP3 promoter with no internal modifications was prepared by annealing and extending 85-mer oligonucleotide primers (Integrated DNA Technologies, Coralville, IA) with a 20-nt overlap ranging from nucleotides 1,471,577 - 1,471,726 in the *M. tuberculosis* H37Rv genome [69] and HPLC purified as previously described [64]. RbpA (saturating concentration), *M. tuberculosis* RNAP-σ^A holoenzyme and linear dsDNA template were incubated at 37ºC for 10mins. Reactions were initiated by adding 2.5 µl of a substrate mixture containing 200 GpU, UTP, and ^{32}P radiolabeled UTP and incubating at 37ºC for 10 minutes to allow for production of a 3 nt product in 20 µl reactions that included a final concentration of 2 µM RbpA, 100 nM *M. tuberculosis* RNAP-σ^A holoenzyme, 10 nM linear dsDNA template, 1 mM DTT, 0.1 mg/ml BSA (NEB), 200 µM GpU, 20 µM UTP, 0.2 µl of ^{32}P radiolabeled UTP, 75 mM NaCl, 10.1 mM $MgCl_2$, 2 µM $ZnCl_2$, 18 mM Tris pH 8.0, 0.01 mM EDTA, 5% glycerol and 0.1 mM β-mercaptoethanol. Reactions were stopped with 2X formamide stop buffer (98% [vol/vol] formamide, 5 mM EDTA and 0.05% w/v bromophenol blue). Reaction products were resolved by 22% polyacrylamide-urea gel electrophoresis and exposure to autoradiography film. Products were quantified using ImageJ.

Multi-round *in vitro* transcription. RbpA , linear DNA PCR-amplified from the pMSG434 plasmid containing the *M. tuberculosis rrn*AP3 promoter including positions -39 to +4, and NTPs (2 mM ATP, 2 mM CTP, 0.1 mM UTP, 4 mM GTP and 0.2 µl per reaction of ^{32}P radiolabeled GTP) were mixed to a final volume of 23 µl. Reactions were initiated with the

addition of *M. tuberculosis* RNAP-σ^A holoenzyme and incubated for 1 hour at 37ºC. The reactions included a final concentration of 2 μM RbpA, 0.8 nM of template DNA, 40 nM *M. tuberculosis* RNAP-σ^A holoenzyme, 1 mM DTT, 0.1 μg/ml BSA, 200 μM ATP, 200 μM CTP, 10 μM UTP, 400 μM GTP, 0.2 μl of ^{32}P labeled GTP, 75 mM NaCl, 10.1 mM $MgCl_2$, 2 μM $ZnCl_2$, 18 mM Tris pH 8.0, 0.01 mM EDTA, 5% glycerol and 0.1 mM β-mercaptoethanol. Reactions were stopped with 2X formamide stop buffer (98% [vol/vol] formamide, 5 mM EDTA and 0.05% w/v bromophenol blue). Reaction products were resolved by 10% polyacrylamide-urea gel electrophoresis, exposure to phosphoimager screen and imaged using an Amersham Typhoon scanner.

Fidaxomicin Dose Response Curve. RbpA diluted in RbpA storage buffer, *M. tuberculosis* RNAP-σ^A holoenzyme diluted in holoenzyme storage buffer and 10X transcription buffer were mixed to a final volume of 14.5 μl. Fidaxomicin diluted in DMSO to a concentration of 2000 μM, 200 μM, 20 μM, 2 μM and 0.2 μM or DMSO was added to the reaction mixtures and the reaction mixtures were incubated at 37ºC for 10mins. The linear dsDNA template containing the *M. tuberculosis rrn*AP3 promoter as described in the 3-nucleotide *in vitro* transcription section, was added to reaction mixtures and incubated at 37ºC for 15mins. Reactions were initiated by adding GpU dinucleotide, UTP and ^{32}P radiolabeled UTP to allow for production of a 3 nt product in a 20 μl mixtures including 2 μM RbpA, 100 nM *M. tuberculosis* RNAP-σ^A holoenzyme, 10 nM linear dsDNA template, 1mM DTT, 0.1% mg/ml BSA (NEB), 200 μM GpU, 20 μM UTP, 0.2 μl of ^{32}P radiolabeled UTP, 75 mM NaCl, 10.1 mM $MgCl_2$, 2 μM $ZnCl_2$, 18 mM Tris pH 8.0, 0.01 mM EDTA, 5% glycerol, 0.1 mM β-mercaptoethanol and either 100 μM, 10 μM, 1.0 μM, 0.1 μM, 0.01 μM or 0 μM of fidaxomicin. Reactions were stopped with 2X formamide stop buffer and resolved by 22% polyacrylamide-urea gel electrophoresis and

exposure to autoradiography film. Products were quantified using ImageJ and IC50 values were calculated using Prism software by four parameter normalized (maximum = 100% and minimum = 0%) nonlinear regression fitting using logarithmic normalization of the Fdx concentrations included in these experiments.

Results

RbpA E17 and R10 synergize to promote Fdx activity against *M. tuberculosis* RNAP-σ^A *in vitro*

In vitro assays that monitor the production of a 3 nucleotide (nt) product as a proxy of RP_o stability have shown that addition of Fdx to *M. tuberculosis* RNAP-σ^A holoenzymes reduces the amount of RP_o formed following the subsequent addition of NTPs and a DNA template harboring the *M. tuberculosis rrnA*P3 promoter [67]**(Figure 1a and 1b)**. Addition of wild-type (WT) $RbpA_{Mtb}$ ($RbpA_{Mtb}^{WT}$) to the RNAP-σ^A holoenzymes increases the sensitivity of the RNAP-σ^A holoenzyme to Fdx in this assay, and this is dependent on the presence of the NTT (amino acids 1-25 in $RbpA_{Mtb}$, deleted in the $RbpA_{Mtb}^{26\text{-}111}$ mutant)[67]**(Figure 1a and 1b)**. Structural studies predicted that the NTT contributes contacts with Fdx when the antibiotic is bound to the *M. tuberculosis* RNAP-σ^A holoenzyme (PDB: 6BZO), specifically through a water mediated interaction between RbpA E17 and Fdx **(Figure 1c)**[67]. To determine whether the predicted interaction between Fdx and RbpA E17 underpins NTT-dependent Fdx activity, we calculated the concentration of Fdx to inhibit 50% of RP_o (IC50) formed by *M. tuberculosis* RNAP-σ^A on the *rrnA*P3 promoter in the presence of $RbpA_{Mtb}^{WT}$ versus an $RbpA_{Mtb}^{E17A}$ mutant protein. The activity of Fdx against the *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}^{E17A}$ was nearly equal to Fdx activity against the *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}^{WT}$, indicating that the contribution of RbpA E17 to the interaction between Fdx and

RNAP-σ^A is not required for Fdx activity against the *M. tuberculosis* RNAP-σ^A **(Figure 1a and 1b)**.

The structure in Boyaci *et al*. also highlights potential van der Waals interactions between RbpA R10 and Fdx in the RNAP-σ^A holoenzyme bound to double stranded forked DNA (PDB: 6BZO)[67]**(Figure 1c)**, however, given the distance between RbpA R10 and Fdx, one would predict this to be a weak interaction. In a separate structure of RbpA bound to *M. tuberculosis* RNAP-σ^A in complex with two double stranded forked DNA molecules that mimics the RP_o (PDB: 6C04), the RbpA R10 positively charged side chain is positioned within 2.4 angstroms of the negatively charged side chain of $\sigma^A{}_{3.2}$ D441, forming a polar interaction [67]**(Figure 1d)**. Fdx activity against *E. coli* RNAP-σ^{70} holoenzyme lacking $\sigma^{70}{}_{3.2}$ is attenuated approximately 20-fold [134], indicating the $\sigma^{70}{}_{3.2}$ contributes to Fdx inhibition of the *E. coli* RNAP and if RbpA R10 interacts with $\sigma^{70}{}_{3.2}$, this may also affect Fdx activity. To examine whether RbpA R10 contributes to *M. tuberculosis* RNAP-σ^A Fdx sensitivity, we measured Fdx IC50s against the *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}{}^{R10A}$. Similar to the $RbpA_{Mtb}{}^{E17A}$ mutant, we observed no change in Fdx IC50s against the *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}{}^{R10A}$ compared to $RbpA_{Mtb}{}^{WT}$ **(Figure 1a and 1b)**, indicating that the R10 residue is not required for RbpA NTT dependent Fdx activity. To determine the effect of disrupting the contacts made by the both RbpA E17 and R10, we measured the Fdx IC50 against *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}{}^{R10A/E17A}$. Mutating both the R10 and E17 residues resulted in an approximately three-fold increase in the Fdx IC50 compared to $RbpA_{Mtb}{}^{WT}$ **(Figure 1a and 1b)**, indicating that loss of one of these residues increases the importance of the other for Fdx activity. However, the IC50 of Fdx in the presence of $RbpA_{Mtb}{}^{R10A/E17A}$ is still five-fold lower than that of $RbpA_{Mtb}{}^{26\text{-}111}$, the ΔNTT mutant, and not

significantly different compared to $RbpA_{Mtb}^{WT}$, suggesting that additional mechanisms contribute to NTT-dependent Fdx activity.

The RbpA CD and conserved residues in the BL and SID do not affect Fdx activity against the *M. tuberculosis* RNAP-σ^A *in vitro*

To further investigate the mechanism of RbpA NTT-dependent Fdx activity, we identified conserved residues within the *M. tuberculosis* RbpA NTT (R4, L6, R7, and S15) that were positioned near various functional domains in the RNAP-σ^A holoenzyme to mutagenize and monitor for effects on Fdx activity *in vitro* (PDB: 6BZO)**(Figure 2a and 2b)**. RbpA R4 is positioned near the RNAP β Sw3, RbpA L6 is positioned near the RNAP β' lid, RbpA R7 is positioned near $\sigma^A_{3.2}$, and RbpA S15 is positioned near the RNAP β' ZBD. We evaluated the role of the positively charged side chains of R4 and R7 by substituting these amino acids with either alanine or glutamate, assessed the length of the hydrophobic L6 side chain by substituting the leucine with alanine, and determined the role of the polar S15 side chain by substituting the serine with alanine. Addition of any of these RbpA mutants to the *in vitro* assays for Fdx inhibition of RP_o resulted in nearly identical IC50s as observed in the presence of $RbpA_{Mtb}^{WT}$ **(Figure 2c)**, demonstrating that none of these residues are required for NTT-dependent Fdx activity *in vitro*.

Until this point, the NTT is the only RbpA domain that has been investigated for effects on Fdx activity. Although the RbpA CD, BL, and SID are not predicted to directly interact with Fdx, we reasoned that RbpA binding to the RNAP-σ^A holoenzyme could impact Fdx activity through an indirect mechanism, such as altering RNAP clamp dynamics. Fdx only binds to the open clamp conformation that exists in the RNAP closed promoter complex (RP_c) and locks the RNAP in the open clamp conformation. The closing of the clamp that occurs during RP_o restricts the volume of the Fdx binding cleft, making Fdx unable to bind RP_o. To determine whether the

RbpA CD affects Fdx activity against the *M. tuberculosis* RNAP-σ^A, we measured the IC50 of Fdx inhibition of *M. tuberculosis* RNAP-σ^A RP_o in the presence of $RbpA_{Mtb}^{72\text{-}111}$, which lacks both RbpA NTT and CD. Deletion of the RbpA NTT and CD resulted in an IC50 similar to a deletion of the NTT alone **(Figure 1b and 2c)**, indicating that the presence of the RbpA CD does not affect Fdx activity against *M. tuberculosis* RNAP-σ^A *in vitro*. We then determined whether the RbpA BL and SID play a role in Fdx activity against the *M. tuberculosis* RNAP-σ^A *in vitro* by measuring Fdx activity against *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}^{R79A}$ or $RbpA_{Mtb}^{R88A}$, mutations that have been shown to weaken the activities of the BL or SID, respectively [22,106,125]. Fdx activity against *M. tuberculosis* RNAP-σ^A in the presence of $RbpA_{Mtb}^{R79A}$ or $RbpA_{Mtb}^{R88A}$ was similar to reactions containing $RbpA_{Mtb}^{WT}$, demonstrating that mutating residues critical for RbpA BL and SID activity does not affect Fdx activity *in vitro* in this assay **(Figure 1b and 2d)**.

Multiple RbpA domains and CarD impact Fdx activity *in vivo*

Previous work showed that truncation of the RbpA NTT decreases Fdx activity both *in vitro* and *in vivo* in *Mycobacterium smegmatis*[67]. Based on the *in vitro* results above, we expected that only *M. smegmatis* expressing $RbpA_{Mtb}^{26\text{-}111}$ and $RbpA_{Mtb}^{72\text{-}111}$ would be more resistant to Fdx than *M. smegmatis* expressing $RbpA_{Mtb}^{WT}$. To investigate the effect of mutations in RbpA on Fdx sensitivity *in vivo*, we used a strain we previously engineered that expresses $rbpA_{Mtb}^{WT}$ at the *attB* site of *M. smegmatis* and has the endogenous *rbpA* gene deleted [125]. We then attempted to replace the $rbpA_{Mtb}^{WT}$ gene at the *attB* site in *M. smegmatis* with alleles encoding each of the RbpA mutants studied in Figures 1 and 2 using a gene swapping method [95,111,125]. We have previously used this approach to generate a *M. smegmatis* strain expressing $rbpA_{Mtb}^{72\text{-}111}$, which has a deletion of both the NTT and CD, as its only *rbpA* allele [125]. However,

we were unable to generate a viable strain expressing *rbpA*$_{Mtb}^{26\text{-}111}$, which deletes only the NTT, in place of *rbpA*$_{Mtb}^{WT}$. In contrast, we were able to replace the *rbpA*$_{Mtb}^{WT}$ allele with the *M. smegmatis* allele *rbpA*$_{Msm}^{28\text{-}114}$, which has previously been used to study the NTT in *M. smegmatis* [67,106]. Similar to our previous report with the *M. smegmatis* strain expressing RbpA$_{Mtb}^{72\text{-}111\ 125}$, RbpA$_{Msm}^{28\text{-}114}$ and RbpA$_{Msm}^{72\text{-}114}$ strains also exhibited a slow growth phenotype **(Figure 3a)**, confirming that while the NTT and CD are not required for viability in *M. smegmatis*, they are important domains for RbpA activity. Using the gene swapping approach, we found that all of the RbpA NTT point mutants could support viability in *M. smegmatis*, although strains expressing RbpA$_{Mtb}^{L6A}$, RbpA$_{Mtb}^{R7A}$, RbpA$_{Mtb}^{R7E}$ and RbpA$_{Mtb}^{S15A}$ exhibited significantly decreased growth rate compared to RbpA$_{Mtb}^{WT}$ in LB media **(Figure 3a)**. The slower growth rates in these RbpA point mutants suggests that the residues targeted may contribute to NTT activity in *M. smegmatis*.

To examine the Fdx sensitivity of each *M. smegmatis* strain, we used a zone of inhibition assay, similar to previous studies [67,94]. By spreading approximately 2.5×10^8 colony forming units (CFUs) of bacteria on an agar plate and spotting 10 µl of 100, 250, or 500 µM Fdx dissolved in dimethyl sulfoxide (DMSO) onto a disk placed onto the plate, the bacteria form a lawn after incubation at 37^{o}C for 2 days and a zone absent of bacterial growth indicates growth inhibition by Fdx. DMSO had no effect on *M. smegmatis* growth in this assay and did not generate a zone of clearing on its own, whereas incubation of *M. smegmatis* with Fdx resulted in growth inhibition **(Figure 3b)**. We compared the radii of the zones of inhibition formed on each *M. smegmatis* mutant with Fdx and reproduced previous findings that deletion of the RbpA NTT results in resistance to Fdx *in vivo* (RbpA$_{Mtb}^{72\text{-}111}$, RbpA$_{Msm}^{28\text{-}114}$ and RbpA$_{Msm}^{72\text{-}114}$ mutants in **(Figure 3b and 3c)**, which is consistent with the *in vitro* findings **(Figure 1b and Figure 2c)**. In

contrast, the RbpA$_{Mtb}^{R10A/E17A}$ mutant was not more resistant to Fdx *in vivo*, which is discordant from the *in vitro* findings where we observe a non-significant but trending decrease in RbpA$_{Mtb}^{R10A/E17A}$ sensitivity to Fdx compared to RbpA$_{Mtb}^{R10A/E17A}$ **(Figure 1b, Figure 3c and supporting information 1)**. In addition, the RbpA$_{Mtb}^{R4A}$ and RbpA$_{Mtb}^{R7A}$ mutants showed increased sensitivity to Fdx at all concentrations tested and the RbpA$_{Mtb}^{R4E}$ mutant displayed increased resistance when 100 μM Fdx was spotted on the disks, despite these mutants exhibiting no change in Fdx IC50 *in vitro* **(Figure 2c and Figure 3b and 3c)**. The rest of the NTT mutations did not affect Fdx sensitivity in *M. smegmatis* **(Figure 3c and supporting information 1)**. Strikingly, the *M. smegmatis* RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ mutants, which have decreased affinity for DNA and the σ factor, respectively, were significantly more sensitive to Fdx treatment **(Figure 3b and 3c)**. These *in vivo* data highlight other contributors to RbpA's effect on Fdx activity that exist in the bacteria but are not recapitulated in the *in vitro* assay.

RbpA stabilizes RNAP-σ^A RP$_o$ *in vitro*, which requires binding of the SID to the RNAP-σ^A holoenzyme and binding of the BL to the DNA. In contrast, the NTT antagonizes RbpA's RP$_o$ stabilizing activity. Accordingly, RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ decrease RP$_o$ stability and RbpA$_{Mtb}^{26\text{-}111}$ increases RP$_o$ stability *in vitro*. These effects on RP$_o$ stability follow the same pattern as observed with Fdx sensitivity *in vivo*, where the RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ mutants are more sensitive, and the RbpA$_{Msm}^{28\text{-}114}$ mutant is more resistant **(Figure 3b and 3c)**. RP$_o$ formation involves closing of the RNAP clamp, which is predicted to block binding of Fdx to the RNAP and could explain this pattern of Fdx sensitivity observed in the RbpA mutants *in vivo*. CarD is another essential transcription factor in mycobacteria that functions to stabilize RP$_o$. We reasoned that if RP$_o$ stability was linked to Fdx activity *in vivo*, then *M. smegmatis* strains expressing the CarD$_{Mtb}^{R25E}$ mutant that has a weaker affinity for the RNAP and is defective in

stabilizing RP_o would be more sensitive to Fdx than *M. smegmatis* expressing $CarD_{Mtb}^{WT}$. Indeed, when we performed the zone of inhibition assays on these strains, we found that the R25E mutation in CarD also increased the sensitivity of *M. smegmatis* to Fdx **(Figure 3d and 3e)**, possibly in part due to a contribution of RP_o stability to Fdx activity *in vivo*. In addition, Fdx has been shown to decrease the affinity of CarD to RNAP *in vitro*, which may also contribute to the increased susceptibility of the $CarD_{Mtb}^{R25E}$ mutant.

Mutation of residues within the RbpA NTT positioned near the RNAP β' lid, RNAP $\sigma^{A}_{3.2}$, and the RNAP ZBD increase RP_o stabilization by RbpA

In contrast to the RbpA BL and SID domains that promote RP_o stability *in vitro*, the RbpA NTT and CD antagonize RP_o stability [125], but it is unknown how the NTT and CD impart this effect on the RNAP-σ^A holoenzyme. To gain insight into this area, we examined how the conserved residues within the *M. tuberculosis* RbpA NTT, contributed to the effect of the NTT on RP_o stability in the 3 nt transcription assay. As previously reported [125], addition of $RbpA_{Mtb}^{26\text{-}111}$ or $RbpA_{Mtb}^{72\text{-}111}$ to *M. tuberculosis* RNAP-σ^A and the *rrnA*P3 promoter increases RP_o stability compared to $RbpA_{Mtb}^{WT}$, although the $RbpA_{Mtb}^{26\text{-}111}$ did not reach statistical significance in this assay **(Figure 4a and 4b)**. Alanine substitutions in RbpA L6, predicted to be positioned near hydrophobic residues in the RNAP β' lid, or RbpA R7, positioned near $\sigma^{A}_{3.2}$, resulted in a higher level of RP_o stability compared to $RbpA_{Mtb}^{WT}$ **(Figure 4a and 4b)**. In addition, mutation of RbpA S15 or E17, two residues positioned near the RNAP β' ZBD, also resulted in increased RP_o stability compared to $RbpA_{Mtb}^{WT}$. Conversely, glutamate substitution for RbpA R4, positioned near RNAP β Sw3, resulted in a lower level of RP_o stability compared to $RbpA_{Mtb}^{WT}$. These data suggest that the NTT could antagonize RNAP-σ^A RP_o stability on the *rrn*AP3 promoter through contacts with RNAP β' lid, $\sigma^{A}_{3.2}$ and β' ZBD while also making other

interactions with RNAP β Sw3 that support RNAP-σ^A RP_o stability. Notably, although the mutations in RbpA L6, R7, S15, and E17 resulted in increased RP_o, these mutations did not lead to decreased sensitivity to Fdx, uncoupling the effects of the RbpA NTT on RP_o stability and Fdx sensitivity **(Figure 4c)**.

The RbpA NTT promotes *M. tuberculosis* RNAP-σ^A full-length transcription from the *M. tuberculosis rrn*AP3 promoter

RbpA is essential in mycobacteria and required to stabilize RNAP-σ^A and RNAP-σ^B RP_o *in vitro* [107–109,125]. Assuming that the ability to stabilize RP_o is linked to RbpA's essential role in mycobacteria, it remains unknown how mycobacteria would benefit from RbpA NTT antagonism of RP_o, but suggests that there is a significant consequence of NTT activity on transcription. To determine the impact of the NTT on full length transcript production, we performed multi-round *in vitro* transcription assays comparing *M. tuberculosis* RNAP-σ^A full-length transcript production from the *M. tuberculosis rrn*AP3 in the presence of either $RbpA_{Mtb}^{WT}$ or $RbpA_{Mtb}^{26\text{-}111}$. Similar to previous reports [108], we found that addition of $RbpA_{Mtb}^{WT}$ resulted in increased full length transcript production from *rrnA*P3 **(Figure 5)**. Deletion of the NTT attenuated full length transcript production **(Figure 5)**, despite increasing RP_o stability compared to $RbpA_{Mtb}^{WT}$ **(Figure 4)**. These data suggest that the NTT destabilization of RP_o facilitates transition of the RNAP through intermediate states that we have previously identified during transcription initiation [69].

Discussion

Prior studies on RbpA have focused almost exclusively on the SID interaction with σ factor and the BL interaction with DNA, leaving the NTT and CD largely uncharacterized. Structural studies have provided tremendous insight into the potential interactions between the

NTT and CD with multiple RNAP-σ^A holoenzyme domains[10,67,106]. Herein we probe the roles for these interactions in RbpA activity through directed genetic mutation of conserved residues within the NTT. We test the prediction that RbpA R10 and E17 contribute contacts with the antibiotic Fdx that are important for its activity against *M. tuberculosis* RNAP-σ^A. We find that *in vitro*, mutation of both residues affects Fdx activity against the *M. tuberculosis* RNAP-σ^A **(Figure 1a and 1b)**, however, it is still not clear whether RbpA R10 and E17 promote RbpA NTT-dependent Fdx activity through direct interaction with Fdx or through an alternative mechanism. Maintenance of partial Fdx activity against *M. tuberculosis* RNAP-σ^A bound by $RbpA_{Mtb}^{R10A/E17A}$ *in vitro* indicates that additional RbpA NTT residues or perhaps the entire structural domain mediate RbpA NTT-dependent Fdx activity. *In vivo*, the $RbpA_{Mtb}^{R10A/E17A}$ mutant did not affect Fdx sensitivity in *M. smegmatis*, suggesting that RbpA impacts Fdx activity independent of its contacts with the antibiotic. In support of this interpretation, we identify several RbpA NTT point mutations that alter *M. smegmatis* sensitivity to Fdx *in vivo* **(Figure 3b and 3c)** despite being in residues not expected to interact directly with Fdx. In addition, the substitution of R88A that weakens RbpA interaction with the RNAP *in vivo* [125], and thus would be expected to decrease *M. smegmatis* sensitivity to Fdx since less RbpA would be associated with RNAP-σ^A, significantly increased *M. smegmatis* sensitivity to Fdx. This suggests that the loss of RbpA activity conferred by the R88A substitution has a greater impact on Fdx sensitivity than the weakened $RbpA_{Mtb}^{R88A}$-RNAP interaction. The experiments herein shed new light on the complex mechanism of RbpA mediated Fdx activity.

Although it is unclear how the different RbpA mutants alter *M. smegmatis* sensitivity to Fdx, we provide evidence that RP_o stability itself may have some effect on Fdx sensitivity *in vivo* where $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ and $CarD_{Mtb}^{R25E}$ mutants that decrease RP_o stability increase *M.*

smegmatis sensitivity to Fdx. Similarly, truncations that increase RbpA RP_o stabilizing activity, $RbpA_{Mtb}^{26\text{-}111}$ and $RbpA_{Mtb}^{72\text{-}111}$, decrease *M. smegmatis* sensitivity to Fdx. However, analysis of RbpA NTT point mutants yielded no correlation between RP_o stability and *M. smegmatis* sensitivity to Fdx, making it unclear what role the equilibrium between RP_c and RP_o plays in determining *M. smegmatis* sensitivity to Fdx.

Our assays revealed differences in the effects of RbpA mutants on Fdx sensitivity *in vitro* compared to *in vivo*. These discrepancies may be due in part to the limited scope of the *in vitro* assay used here and in previous studies to probe Fdx activity [67], where Fdx is added to RbpA and RNAP-σ^A holoenzyme in the absence of DNA. Whereas in the cell, RNAP-σ^A holoenzyme could be bound to DNA prior to Fdx binding. This limitation may bias the *in vitro* assay towards identifying the factors that affect Fdx binding to free RNAP-σ^A holoenzyme complex. As such, RbpA NTT substitutions that impact *M. smegmatis* sensitivity to Fdx but show no effect *in vitro* ($RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$ and $RbpA_{Mtb}^{R7A}$) could result from these RbpA NTT residues being important for Fdx binding to DNA bound holoenzyme rather than free RNAP. This notion is supported by a comparison of RbpA NTT interactions with Fdx bound to free *M. tuberculosis* RNAP (PDB: 6C06) and Fdx bound to *M. tuberculosis* RNAP in complex with DNA (PDB: 6BZO)[67]. In the absence of DNA, the RbpA NTT is positioned closer to RNAP β' ZBD, RNAP β' lid and RNAP $\sigma^A{}_{3.2}$ in what could be described as a more compact structure compared to the DNA bound complex. The more compact structure results in several polar interactions between RbpA and the RNAP-σ^A holoenzyme that do not occur in the presence of DNA. The differences between how RbpA interacts with DNA bound versus unbound RNAP-σ^A holoenzyme highlight the need to biochemically characterize Fdx activity against not just free RNAP but the additional RNAP complexes that exist within the bacteria. The *in vitro* assay also excludes other RNAP

holoenzymes and RNAP interacting proteins present in the bacteria that may affect Fdx activity *in vivo*. In addition to these direct effects on RNAP, truncation of the RbpA NTT results in global dysregulation of gene expression in *M. smegmatis* [106,125], which could also affect sensitivity to Fdx.

RbpA has been shown to increase RP_o stability and based on recent modeling [135], we hypothesize that this regulatory mechanism can be capable of transcriptional activation and repression. In this study, we find that RbpA increases full length transcript production from the *M. tuberculosis rrnA*P3 promoter. The RbpA NTT antagonizes the ability of RbpA to stabilize *M. tuberculosis* RNAP-σ^A RP_o while decreasing at least one rate constant in the RP_c to RP_o transition [125]. At the *rrnA*P3 promoter, NTT antagonism of RP_o stability results in increased full length transcript production. Mechanistically, one of the ways that RbpA NTT antagonism of RP_o stability may result in an increase in transcriptional flux is through enhancing the rate of promoter escape. Loss of RbpA NTT antagonism of RP_o stability may slow the rate of promoter escape and decrease transcriptional flux, though the role of RbpA NTT in promoter escape has not yet been determined [69]. Alternatively, the RbpA NTT could impact one of the intermediate steps between RP_o formation and promoter escape, which we have also demonstrated to be impacted by RbpA *in vitro* [69] Given that the RbpA NTT is predicted to contact multiple functional domains of the RNAP-σ^A holoenzyme that have been shown to affect various steps of transcription initiation (as described in the introduction), the mechanistic basis for NTT regulation of RbpA activity and physiological consequences on the bacteria are likely complex and a fertile area for future investigations.

Acknowledgments

C.L.S. and E.A.G. were supported by grant GM107544 from the National Institutes of Health (NIH). C.L.S. was also supported by a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease Award. J.P. was supported by NIH National Institute of General Medical Sciences grant GM007067 and Stephen I. Morse Graduate Fellowship. D.J. was supported by a Gary K. Ackers Fellowship and an Elliot L. Elson Education and Training Fellowship. D.X.Z. was supported by T32A1007172 from the NIH National Institute of Allergy and Infectious Diseases.

Figures

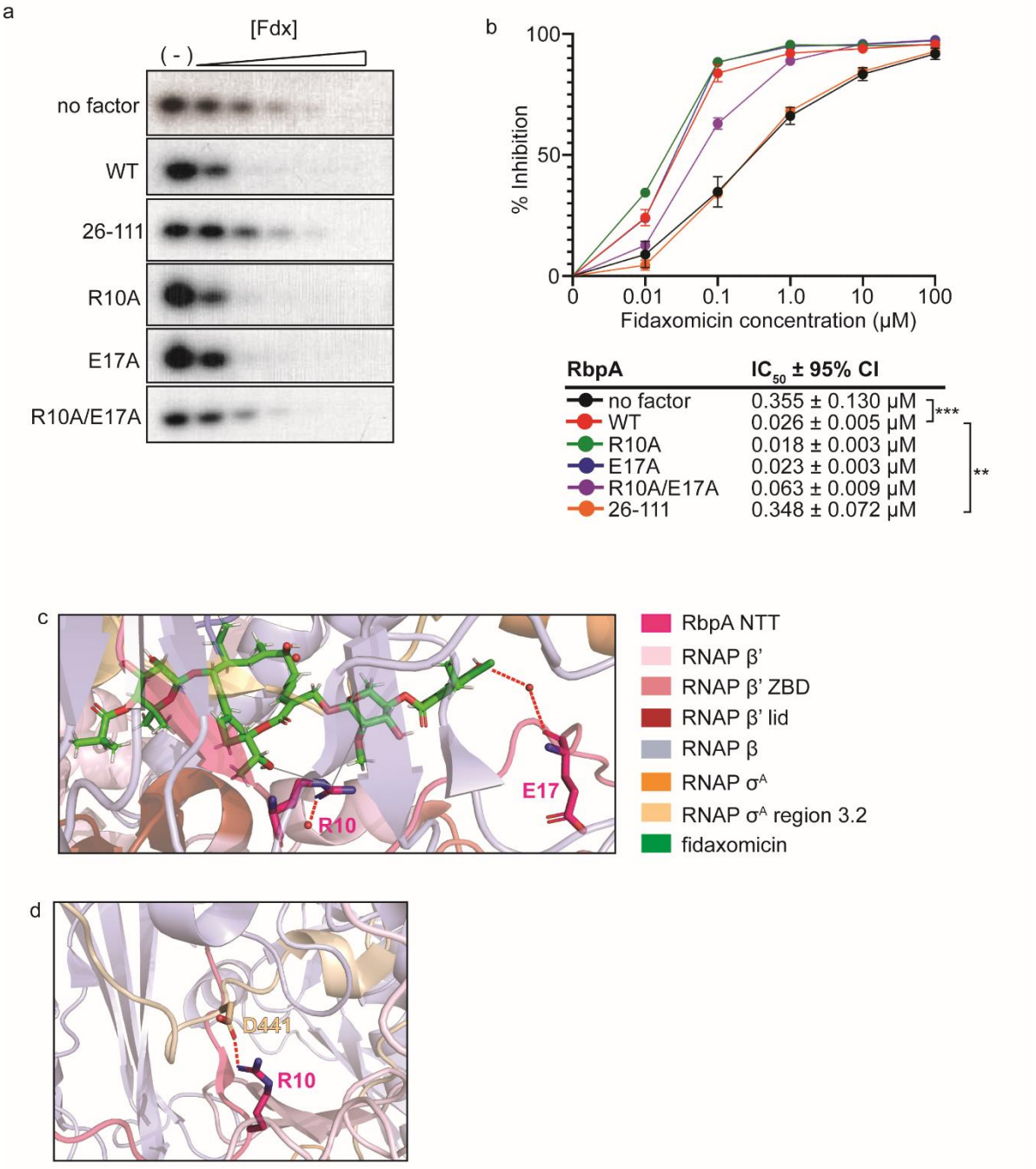

a
[Fdx]
(-)
no factor
WT
26-111
R10A
E17A
R10A/E17A
b
100
50
0
% Inhibition
0
0.01
0.1
1.0
10
100
Fidaxomicin concentration (μM)
RbpA
IC50 ± 95% CI
no factor
0.355 ± 0.130 μM
WT
0.026 ± 0.005 μM
***
R10A
0.018 ± 0.003 μM
E17A
0.023 ± 0.003 μM
**
R10A/E17A
0.063 ± 0.009 μM
26-111
0.348 ± 0.072 μM
c
E17
R10
RbpA NTT
RNAP β'
RNAP β' ZBD
RNAP β' lid
RNAP β
RNAP σA
RNAP σA region 3.2
fidaxomicin
d
D441
R10

Figure 1. RbpA E17 and R10 synergize to promote fidaxomicin inhibition of *M. tuberculosis* RNAP-σ^A holoenzyme activity *in vitro*

a. Representative gels showing Fdx (0 μM, 0.01 μM, 0.1 μM, 1.0 μM, 10 μM, 100 μM) inhibition of *M. tuberculosis* RNAP-σ^A production of 3nt transcripts alone or in complex with $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R10A}$, $RbpA_{Mtb}^{E17A}$, $RbpA_{Mtb}^{R10A/E17A}$, or $RbpA_{Mtb}^{26\text{-}111}$ from a linear dsDNA template containing positions -80 to +70 of *M. tuberculosis rrn*AP3 (relative to the +1 transcription start site).

b. Dose response curves of the results shown in (A). The curves are generated from at least four replicates from at least two different experiments. Percent inhibition at each Fdx concentration included in the plots is the mean +/- SEM. The IC50 for each replicate was calculated by non-linear regression analysis with four parameter fitting of log transformed Fdx concentration versus normalized response, with the mean IC50 and 95% confidence interval listed in the table. Statistical comparisons were performed with one-way analysis of the variance (ANOVA) and Tukey's multiple comparison test. All comparisons to $RbpA_{Mtb}^{WT}$ were included in the analysis but only statistically significant comparisons are indicated in the figure; **, $P < 0.01$; ***, $P < 0.001$.

c. Structural modeling of Fdx binding pocket on the RbpA bound *M. tuberculosis* RNAP-σ^A from PDB structure 6BZO. Fdx and RNAP residues involved in the RNAP-Fdx binding interface are shown with PyMol stick representation while the rest of the structure is shown with PyMol cartoon representation. Polar interactions are indicated by red-dashed lines and potential Van der Waals interactions are shown as gray lines.

d. Structural modeling of RbpA bound *M. tuberculosis* RNAP-σ^A from PDB structure 6C04. RbpA R10 and RNAP σ^A D441 are shown with PyMol stick representation while the rest

of the structure is shown with PyMol cartoon representation. The polar interaction between RbpA R10 and RNAP σ^A D441 is indicated by the red dashed line.

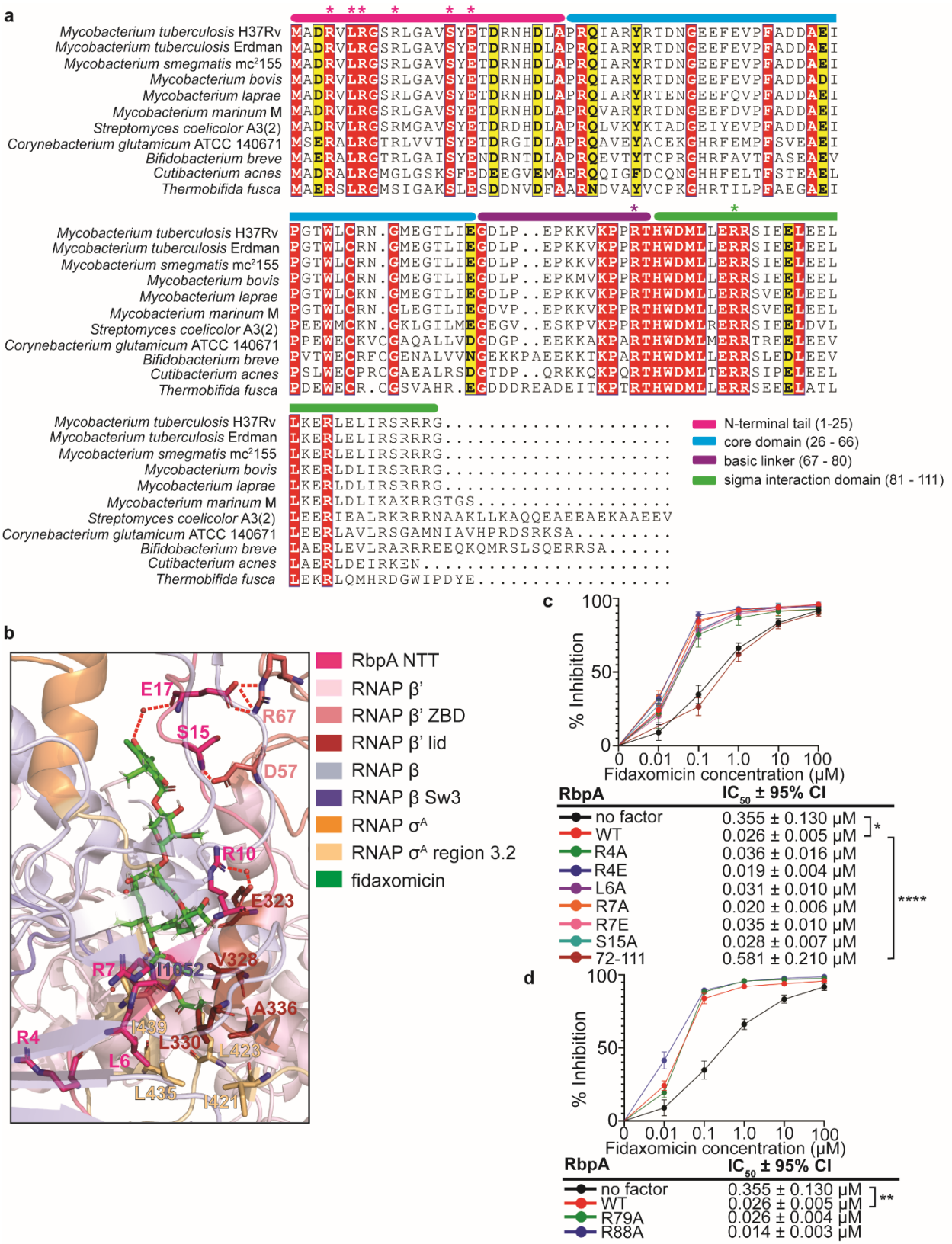

a
Mycobacterium tuberculosis H37Rv
Mycobacterium tuberculosis Erdman
Mycobacterium smegmatis mc²155
Mycobacterium bovis
Mycobacterium laprae
Mycobacterium marinum M
Streptomyces coelicolor A3(2)
Corynebacterium glutamicum ATCC 140671
Bifidobacterium breve
Cutibacterium acnes
Thermobifida fusca
N-terminal tail (1-25)
core domain (26 - 66)
basic linker (67 - 80)
sigma interaction domain (81 - 111)
b
RbpA NTT
RNAP β'
RNAP β' ZBD
RNAP β' lid
RNAP β
RNAP β Sw3
RNAP σA
RNAP σA region 3.2
fidaxomicin
E17
R67
S15
D57
R10
E323
R7
I1052
V328
A336
I439
R4
L330
L423
L6
L435
I421
c
% Inhibition
Fidaxomicin concentration (μM)
RbpA
IC50 ± 95% CI
no factor 0.355 ± 0.130 μM
WT 0.026 ± 0.005 μM
R4A 0.036 ± 0.016 μM
R4E 0.019 ± 0.004 μM
L6A 0.031 ± 0.010 μM
R7A 0.020 ± 0.006 μM
R7E 0.035 ± 0.010 μM
S15A 0.028 ± 0.007 μM
72-111 0.581 ± 0.210 μM
d
% Inhibition
Fidaxomicin concentration (μM)
RbpA
IC50 ± 95% CI
no factor 0.355 ± 0.130 μM
WT 0.026 ± 0.005 μM
R79A 0.026 ± 0.004 μM
R88A 0.014 ± 0.003 μM

Figure 2. The RbpA CD, BL, and SID do not affect fidaxomicin activity against the *M. tuberculosis* RNAP-σ^A *in vitro*

a. Alignment of RbpA. Conserved RbpA NTT residues targeted in this analysis are indicated by an asterisk and the four RbpA structural domains are indicated.

b. Structural modeling of RbpA NTT interactions with the RNAP from PDB structure 6C04. RbpA NTT residues targeted in this analysis are shown with PyMol stick representation while the rest of the structure is shown with PyMol cartoon representation. Zoom in views of areas (a) – (f) show RbpA R4, L6A, R7A, R10, S15 and E17, respectively, and the RNAP residues located near these RbpA NTT residues or predicted to make polar interactions in PyMol are shown in stick representation.

c. and d. Dose response curves of Fdx (0 μM, 0.01 μM, 0.1 μM, 1.0 μM, 10 μM, 100 μM) inhibition of *M. tuberculosis* RNAP-σ^A production of 3nt transcripts alone or in complex with $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7A}$, $RbpA_{Mtb}^{R7E}$, $RbpA_{Mtb}^{S15A}$, $RbpA_{Mtb}^{72\text{-}111}$, $RbpA_{Mtb}^{R79A}$ or $RbpA_{Mtb}^{R88A}$ from linear dsDNA template containing -80 to +70 positions of *M. tuberculosis rrn*AP3 relative to the +1 transcription start site. The curves are generated from at least four replicates from at least two different experiments. Percent inhibition at each Fdx concentration included in the plots is the mean +/- SEM. The WT data from Figure 1 is included in C and D because the experiments were run simultaneously with the same controls. IC50 values and statistics were calculated as done in Figure 1. All comparisons to $RbpA_{Mtb}^{WT}$ were included in the analysis but only statistically significant comparisons are indicated in the figure; *, $P < 0.05$, **, $P < 0.01$; ****, $P < 0.0001$.

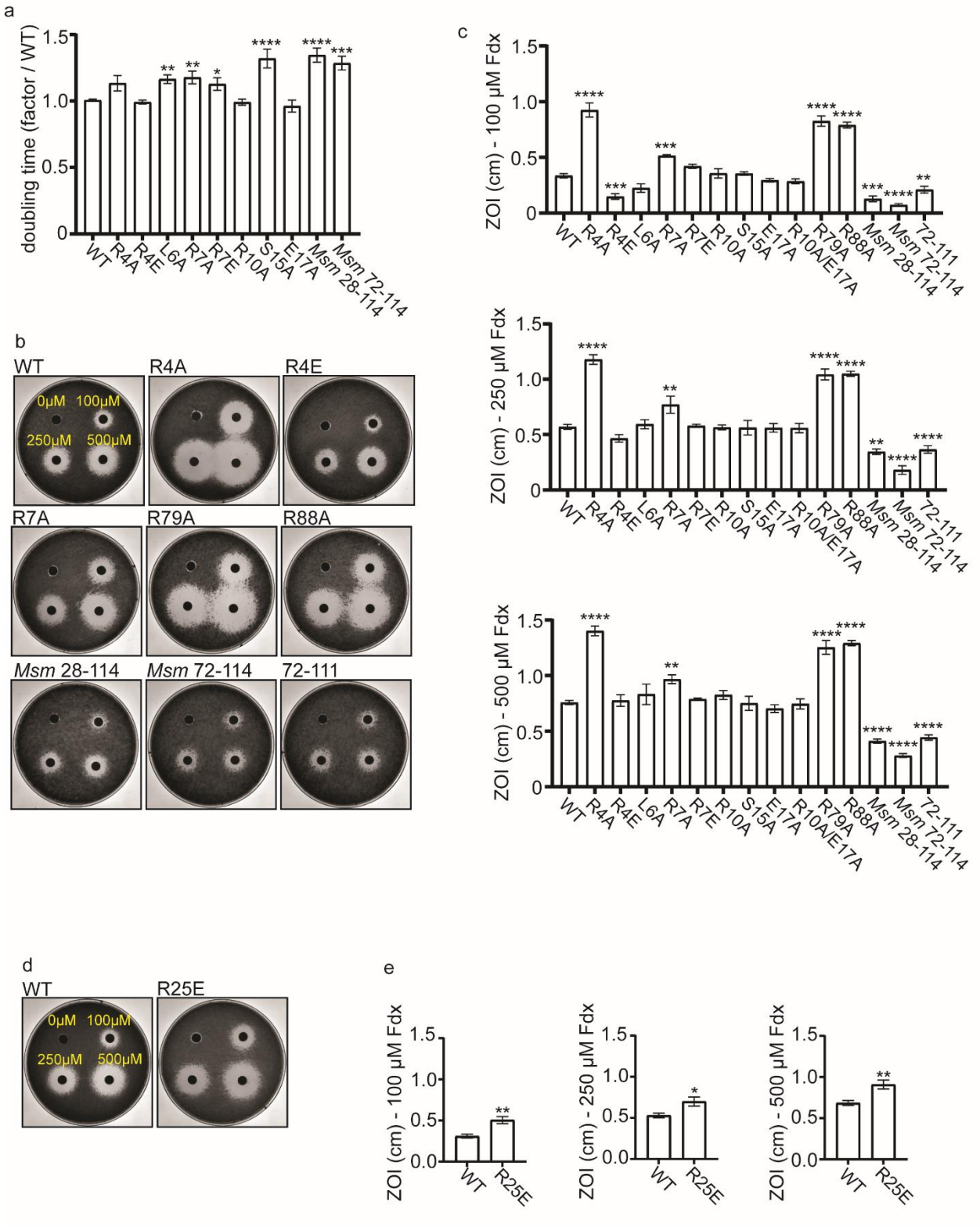

a
doubling time (factor / WT)
1.5
1.0
0.5
0
WT
R4A
R4E
L6A
R7A
R7E
R10A
S15A
E17A
Msm 28-114
Msm 72-114
b
WT
R4A
R4E
0μM
100μM
250μM
500μM
R7A
R79A
R88A
Msm 28-114
Msm 72-114
72-111
c
ZOI (cm) - 100 μM Fdx
ZOI (cm) - 250 μM Fdx
ZOI (cm) - 500 μM Fdx
R10A/E17A
R79A
R88A
72-111
d
R25E
e

Figure 3. Multiple RbpA domains and CarD impact Fdx activity *in vivo*

a. Representative growth curve comparing *M. smegmatis* strains expressing $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7A}$, $RbpA_{Mtb}^{R7E}$, $RbpA_{Mtb}^{R10A}$, $RbpA_{Mtb}^{S15A}$, $RbpA_{Mtb}^{E17A}$, $RbpA_{Msm}^{28\text{-}114}$ or $RbpA_{Msm}^{72\text{-}114}$.

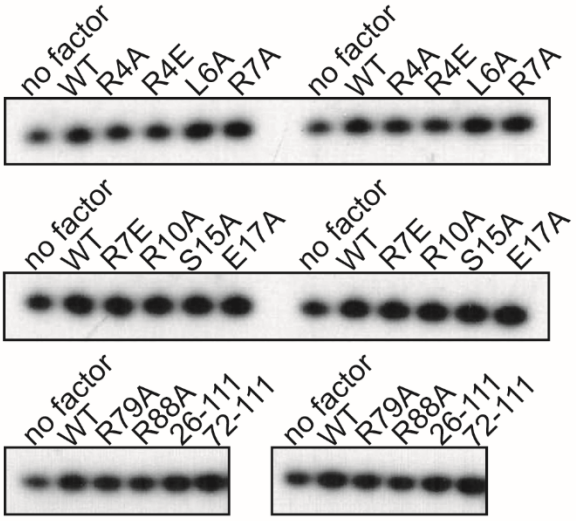
b. Zones of inhibition (ZOI) by Fdx on bacterial lawns of *M. smegmatis* expressing either $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{R7A}$, $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$, $RbpA_{Msm}^{28\text{-}114}$, $RbpA_{Msm}^{72\text{-}114}$, $RbpA_{Mtb}^{72\text{-}111}$.

c. The results are plotted showing the mean ± SEM from at least two experiments with three replicates at 0 μM, 100 μM, 250 μM and 500 μM Fdx. Statistical significance was analyzed by analysis of the variance (ANOVA) and Tukey's multiple comparison test. *, $P < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, not significant. All comparisons were included in the analysis but only statistically significant comparisons are indicated in the figure.

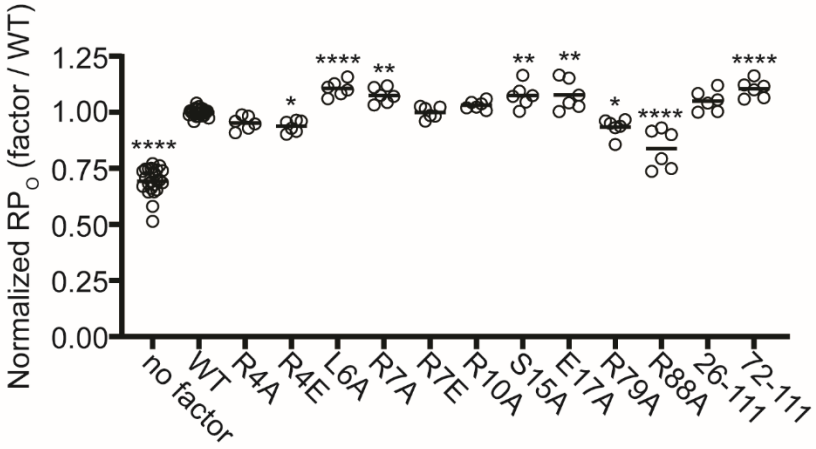
d. ZOI by Fdx on bacterial lawns of *M. smegmatis* expressing either $CarD_{Mtb}^{WT}$ or $CarD_{Mtb}^{R25E}$.

e. The results are plotted showing the mean +/- SEM from three experiments with three replicates at 0 μM, 100 μM, 250 μM and 500 μM Fdx. Statistical significance was analyzed by two-tailed Welch's t test. *, $p < 0.05$; **, $p < 0.01$.

a



b



c

RbpA construct	Fdx IC50	Fdx ZOI	RP$_o$ stability
R4A	NS	more sensitive	NS
R4E	NS	less sensitive	decrease
L6A	NS	NS	increase
R7A	NS	more sensitive	increase
R7E	NS	NS	NS
R10A	NS	NS	NS
S15A	NS	NS	increase
E17A	NS	NS	increase
R10A/E17A	NS	NS	NS
R79A	NS	more sensitive	decrease
R88A	NS	more sensitive	decrease
26-111	increase	less sensitive	NS
72-111	increase	less sensitive	increase

Figure 4. Mutations of residues in the RbpA NTT increases RP_o stability

a. Representative gels of 3nt transcripts produced by *M. tuberculosis* RNAP-σ^A alone or in complex with $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7A}$, $RbpA_{Mtb}^{R7E}$, $RbpA_{Mtb}^{R10A}$, $RbpA_{Mtb}^{S15A}$, $RbpA_{Mtb}^{E17A}$, $RbpA_{Mtb}^{26\text{-}111}$ or $RbpA_{Mtb}^{72\text{-}111}$ from linear dsDNA template containing positions -80 to +70 of *M. tuberculosis rrn*AP3 relative to the +1 transcription start site.

b. Ratio of transcript produced as compared to the average of $RbpA_{Mtb}^{WT}$ replicates included on the same gel. At least two replicates from three different experiments for a total of six replicates were included for each $RbpA_{Mtb}$ variant in the analysis. Results are plotted as individual values with the mean +/- SEM shown. Statistical significance of differences was determined by ANOVA and Tukey's multiple comparison test. *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$. All comparisons to $RbpA_{Mtb}^{WT}$ were included in the analysis, but only statistically significant comparisons are indicated in the figure.

c. Overview of the effect of RbpA mutants on Fdx IC50 (Figs. 1 and 2), Fdx ZOI (Fig. 3), and RP_o stability (Fig. 4) as compared to $RbpA_{Mtb}^{WT}$. NS indicates not significantly different. For Fdx IC50 column; dark green = significantly increased IC50 compared to $RbpA_{Mtb}^{WT}$. For Fdx ZOI column; light green = significantly less sensitive to Fdx compared to $RbpA_{Mtb}^{WT}$ at one Fdx concentration; dark green = significantly less sensitive to Fdx compared to $RbpA_{Mtb}^{WT}$ at more than one Fdx concentration; dark red = significantly more sensitive to Fdx compared to $RbpA_{Mtb}^{WT}$ at more than one Fdx concentration. For RP_o stability column; dark green = significantly increased RP_o compared to $RbpA_{Mtb}^{WT}$; light red = significantly (*, $P < 0.05$) decreased RP_o compared

to RbpA$_{Mtb}^{WT}$; dark red = significantly (****, P < 0.0001) decreased RP_o compared to RbpA$_{Mtb}^{WT}$.

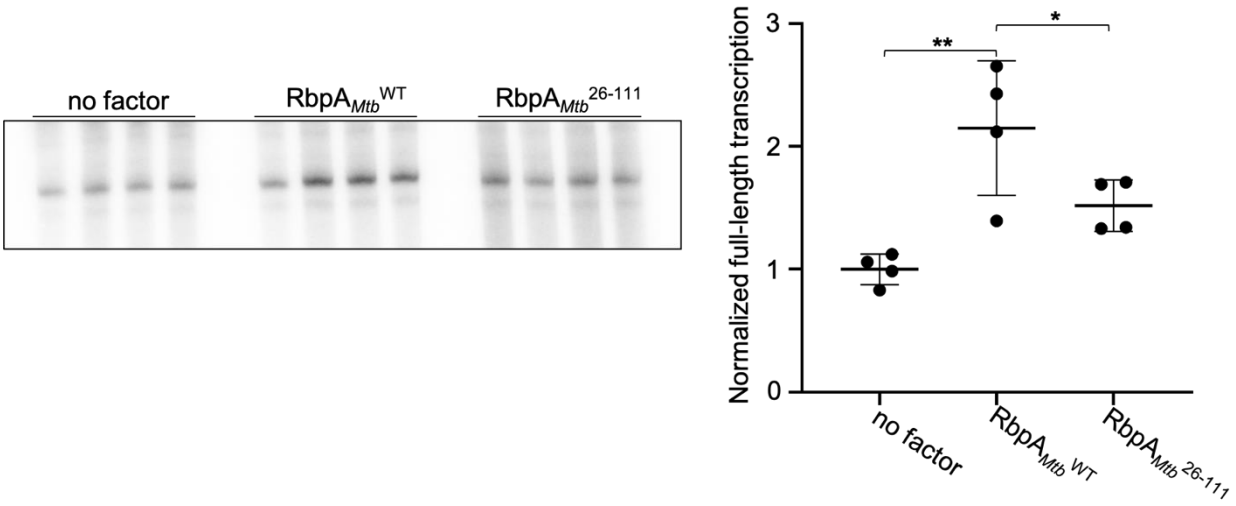
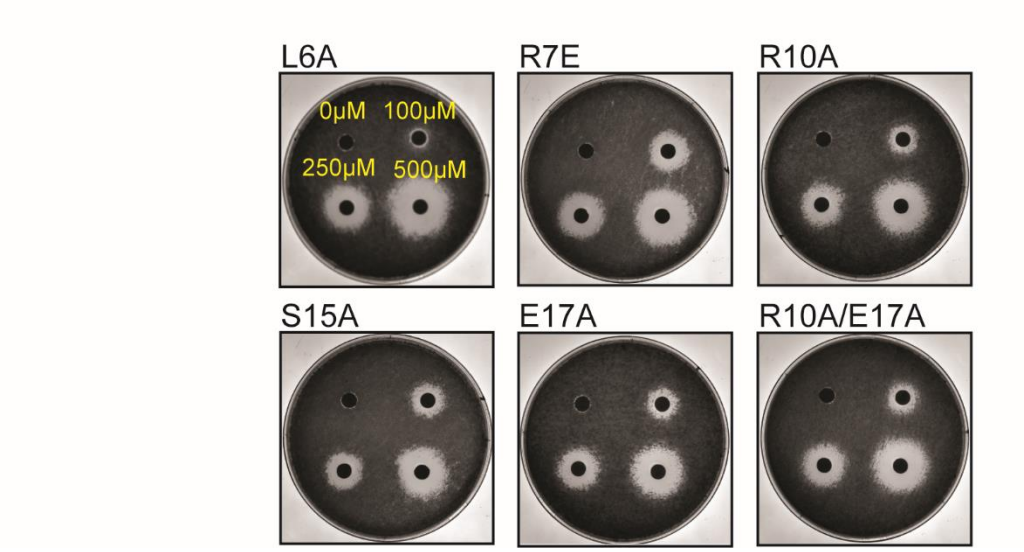


Figure 5. RbpA NTT promotes *M. tuberculosis* RNAP-σ^A full-length transcription on the *M. tuberculosis rrn*AP3 promoter.

a. Gel of full-length transcript production by *M. tuberculosis* RNAP-σ^A alone or in complex with RbpA$_{Mtb}^{WT}$ or RbpA$_{Mtb}^{26-111}$ from the *M. tuberculosis rrn*AP3 promoter on a linear dsDNA template containing positions -39 to +4 of *M. tuberculosis rrn*AP3 relative to the +1 transcription start site.

b. Mean ratios +/- SEM of the amount of transcript produced by each sample as compared to the average of no factor added samples. Statistical significance was analyzed by ANOVA and Dunnett's multiple comparison test. *, $P < 0.05$; **, $P < 0.01$.



Supplemental figure 1. ZOI from 0 μM, 100 μM, 250 μM and 500 μM Fdx on bacterial lawns of *M. smegmatis* expressing $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7E}$, $RbpA_{Mtb}^{R10A}$, $RbpA_{Mtb}^{S15A}$, $RbpA_{Mtb}^{E17A}$ or $RbpA_{Mtb}^{R10A/E17A}$.

Chapter 4: *M. smegmatis* σ^B activates transcription in a RbpA independent manner during logarithmic growth and is made synthetically essential by loss of RbpA R79 and R88 dependent activities

Jerome Prusa, Christina L. Stallings

Abstract

RbpA interacts with *M. tuberculosis*'s essential housekeeping σ^A and the structurally similar but non-essential σ^B. Most studies characterizing σ^B in *M. tuberculosis* have focused on σ^B activities when various stresses are imposed on *M. tuberculosis*. From these studies, it is well understood that σ^B is critical for *M. tuberculosis* stress tolerance. However, σ^B activities in other mycobacterial species during stress or logarithmic growth have not yet been characterized. Furthermore, to what extent RbpA alters σ^B activities under various growth conditions in *M. tuberculosis* or other mycobacteria species is still an open question. *In vitro* analyses of the RbpA-σ^B complex show that RbpA increases σ^B activities at housekeeping promoters, suggesting that RbpA mediates functional overlap between σ^A and σ^B. In this study we interrogate the RbpA dependent and independent σ^B regulon in *M. smegmatis*, a soil dwelling fast growing mycobacterial species, during logarithmic growth and find that the σ^B regulon is almost entirely RbpA independent. Our data also suggests that during logarithmic growth σ^B activates transcription of a cohort of genes that if translated are short and highly charged. We determine that point mutations in RbpA, which *in vitro* blunt RbpA RP_o stabilizing activity and *in vivo* slow *M. smegmatis* growth, alter *sigB* from a non-essential gene to a synthetically essential gene.

Introduction

M. tuberculosis σ^B is a non-essential housekeeping like σ factor and is homologous to *E. coli*'s stationary phase σ factor, σ^S. In *E. coli*, σ^S serves as a general stress response σ factor that accumulates in response to multiple types of stress through a complex regulatory network of small RNAs [136]. One of the defining features of the *E. coli* σ^S general stress response is the

altered expression of gene products seemingly unrelated to whatever stress that triggered the general stress response. The role of σ^B in *M. tuberculosis* is not as well-defined as that of σ^S in *E. coli* but studies that have investigated *sigB* expression or the impact of altering *sigB* expression on *M. tuberculosis* replication during various stresses lend support for σ^B serving as a general stress response σ factor in *M. tuberculosis*. Expression of *sigB* is increased when *M. tuberculosis* enters stationary phase, lacks nutrients during starvation, is exposed to hypoxia, heat or cold temperatures or is treated with thioridozine [137–139]. Deletion of *sigB* increases *M. tuberculosis* sensitivity to hypoxia, cell surface stress, iron starvation and treatment with antibiotics including rifampicin, *p*-aminosalicylic acid, sulfamethoxazole and ethambutol [140–143].

Other studies that have interrogated how *sigB* transcription is regulated also support the notion that σ^B acts a general stress response σ factor. Specifically, *sigB* is expressed from two promoters, the first promoter is activated by the stress-inducible σ factors, σ^E, σ^H and σ^L as well as the MprAB two-component system and the second promoter is regulated by another stress-inducible σ factor, σ^F [144–146]. In conformity with the *E. coli* σ^S paradigm of a general stress response, this data suggests that σ^B protects *M. tuberculosis* from a broad array of stresses rather than a specific stress.

Despite the ample evidence that σ^B is pivotal to *M. tuberculosis* survival during stress, other data suggest that σ^B is not a general stress response σ factor and instead functions more like a housekeeping σ factor (σ^A in *M. tuberculosis* or σ^{70} in *E. coli*). This model is supported by the observation that deletion of *sigB* does not diminish the survival of *M. tuberculosis* in macrophages or murine tissues, which would be expected if σ^B is important for protection against the myriad of stresses imposed on the bacterium during infection [140]. Other evidence suggesting that σ^B is functionally more similar to σ^A or σ^{70} than σ^S is the finding that *sigB* is

expressed at high levels that are comparable to *sigA* during logarithmic growth in *M. tuberculosis*. This is distinct from the expression pattern of *rpoS* (gene nomenclature for σ^S) in *E. coli* where *rpoS* levels are much lower compared to *rpoD* (gene nomenclature for σ^{70}) during logarithmic growth [137,147]. Furthermore, ChIP-seq analysis to define the σ^B regulon in *M. tuberculosis* during logarithmic growth found that σ^B is localized to the promoters of many housekeeping genes [143].

Another feature that differentiates *M. tuberculosis* σ^B from *E. coli* σ^S is σ^B interaction with the essential protein RbpA, which is not encoded by *E. coli*. The housekeeping like behavior of σ^B is enhanced when in complex with RbpA. This was first demonstrated when RbpA was shown to increase σ^B-holoenzyme transcription of the housekeeping *sigA* promoter [108]. How RbpA allows the σ^B-holoenzyme to activate transcription at housekeeping promoters has not yet been fully elucidated but some details have emerged indicating that RbpA allows the σ^B-holoenzyme to overcome a deficiency at housekeeping promoters with a 'minimal' extended -10 element sequence [148,149]. Approximately 70% of *M. tuberculosis*'s promoters have a guanosine at positions -13 or -14 relative to the +1 TSS, which in addition to -11A and -7T within the -10 element, is the minimal housekeeping promoter sequence sufficient for σ^A-holoenzyme activation that can occur without RbpA [149]. In contrast, activity of the σ^B-holoenzyme at this minimal housekeeping promoter requires RbpA [108,149].

Collectively, the work thus far suggests that *M. tuberculosis* σ^B could be important for a stress response program that was not captured in the macrophage or murine infection models while also contributing to the transcription of *M. tuberculosis*'s housekeeping genes in a RbpA dependent manner. However, there are still many questions that need to be answered regarding

σ^B function in *M. tuberculosis* and other mycobacterium species. First, the differences between the role of σ^B in *M. tuberculosis* and σ^S in *E. coli* raise the question of whether there are additional σ^B functions that can be discovered through studying σ^B homologs in other bacterial species. Second, it is unclear whether σ^B only functions in a RbpA dependent manner or whether σ^B also performs RbpA independent functions. Lastly, it is not yet clear whether RbpA-bound σ^B and RbpA-bound σ^A holoenzymes carry out overlapping or distinct functions *in vivo*.

In this study we define the σ^B regulon in *M. smegmatis* during logarithmic growth by examining the transcriptome of *M. smegmatis* Δ*sigB* and determine that this regulon consists of approximately 600 genes. Among these 600 genes, we highlight a novel cohort of transcripts that if translated would be short and highly charged proteins. We compare the transcriptome of *M. smegmatis* Δ*sigB* to a that of a *M. smegmatis* strain with a point mutation in RbpA, R88A, which weakens RbpA binding to σ^A and σ^B, and we find little overlap between the transcriptomes suggesting that the σ^B regulon during logarithmic growth in *M. smegmatis* occurs through a RbpA independent mechanism. We find that *sigB* is synthetically essential in the genetic context of RbpA point mutations including R88A, as well as R79A that weakens RbpA interactions with the DNA phosphate backbone during the RNAP open promoter complex (RP_o).

Experimental Procedures

Media and bacterial strains. All *M. smegmatis* strains were derived from mc^2155 and grown at 37°C in LB medium supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% Tween 80 (broth). The *M. smegmatis* merodiploid strain was constructed by integrating pMSG430-$rbpA_{Mtb}^{WT}$ into the *attB* site of mc^2155. The *M. smegmatis* merodiploid strain was transformed with pDB88, with homology to mc^2155 nucleotides 3928650 to 3929246 and 3929589 to 3930405, to replace the endogenous *rbpA*, using two-step allelic exchange as described

previously [112], thus generating Δ*rbpA*::tet-*rbpA*$_{Mtb}^{WT}$, which was named csm275. Csm275 was transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ to replace the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of the *M. smegmatis* Δ*rbpA attB*::tet-*rbpA*$_{Mtb}^{WT}$ strain using gene swapping as described previously [111]. The transformants were selected with zeocin, and loss of the pMSG430-*rbpA*$_{Mtb}^{WT}$ construct was confirmed by verifying their inability to grow in the presence of kanamycin. The *M. smegmatis* Δ*rbpA*::*tet-rbpA*$_{Mtb}^{WT}$ strain transformed with pDB19-*rbpA*$_{Mtb}^{WT}$ was named csm291. Csm291 was transformed with pMSG430-*rbpA*$_{Mtb}^{R88A}$ to replace the pDB19-*rbpA*$_{Mtb}^{WT}$ construct at the *attB* site of csm291. Csm291 transformed with pMSG430-*rbpA*$_{Mtb}^{R88A}$ was named csm314. Genetic deletion of *sigB* was accomplished through engineering the conditionally replicating phage phAE87 and specialized transduction as previously described [93,150]. *M. smegmatis* Δ*sigB* was named csm375.

Antibiotics and chemicals. In mycobacterial cultures, 20 μg/ml kanamycin, 12.5 μg/ml zeocin or 50 μg/ml hygromycin were used. In *E. coli* cultures, 40 μg/ml kanamycin, 50 μg/ml chloramphenicol, 50 μg/ml streptomycin, and 100 μg/ml ampicillin were used.

RNA-seq analysis.*M. smegmatis* strains csm275, csm314, and csm375 were cultured to an OD_{600} of 0.4 to 0.6, pelleted, resuspended in TRIzol (Thermo Fisher Scientific), and lysed by bead beating (FastPrep; MP Bio, Santa Ana, CA). RNA was extracted with chloroform, precipitated with isopropanol, and resuspended in water. RNA was treated with DNase I (Thermo Fisher Scientific), and RNA integrity and quality were analyzed with an Agilent bioanalyzer. rRNA was removed from samples using the Illumina Ribo-Zero rRNA removal kit. cDNA libraries were generated using an adapted Illumina TruSeq library preparation kit and were quality controlled by analysis of the cDNA size distribution with the Agilent TapeStation. cDNA libraries were pooled and sequenced in a single lane of an Illumina HiSeq 2000 Rapid

Run flow cell with a 50-bp single-end read format. Sequencing reads were demultiplexed and converted to a FASTQ format using Illumina bcl2fastq script. Adapter sequences were trimmed from the raw reads, which were then aligned with the *M. smegmatis* mc^2155 reference genome (GenBank accession number NC_008596) using the STAR aligner [116]. Sequence alignment map (SAM) files generated from alignments were converted to BAM files using SAMTools [117], and aligned reads were counted per genome feature using the BioConductor package Subread featureCounts function [118]. Differential expression analysis and subsequent PCA were performed with BioConductor DESeq2 [119]. Venn diagrams were made with an online tool (https://www.stefanjol.nl/venny). Hypergeometric *P* values and enrichment values were calculated using an online calculator (http://systems.crump.ucla.edu/hypergeometric). The hypergeometric distribution describes the probability of *k* successes in *s* draws, without replacement, from a population of size *N* that contains exactly *M* successes. *N* was defined at the total number of differentially expressed genes in the two RbpA mutant constructs being compared, *s* was defined as the number of differentially upregulated or downregulated genes in one RbpA mutant included in the comparison, *M* was defined as the number of differentially upregulated or downregulated genes in the second RbpA mutant included in the comparison, and *k* was defined as the number of differentially upregulated or downregulated genes shared by the two RbpA mutants in the comparison.

Accession number(s). The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus [120] and are accessible through GEO Series accession number GSE107123.

Results

σ^B regulates transcription in a RbpA independent manner during logarithmic growth

RbpA interacts with the σ^B bound holoenzyme but it is not clear whether all σ^B regulated genes involve RbpA activity. To determine what portion of the σ^B regulon is RbpA dependent we compared the transcriptomes of a *M. smegmatis* strain with a *sigB* deletion (*ΔsigB*) to a *M. smegmatis* strain expressing $RbpA_{Mtb}^{R88A}$, which weakens RbpA's interaction with both σ^A and σ^B. If genes are differentially expressed in both strains, this suggests that these genes are regulated by the RbpA-σ^B complex. Genes that are differentially expressed in only *M. smegmatis ΔsigB* are genes that are regulated by σ^B in a RbpA independent manner. Genes that are differentially expressed in only the $RbpA_{Mtb}^{R88A}$ strain are genes that are regulated by σ^A in a RbpA dependent and σ^B independent manner.

A total of 585 genes are differentially expressed in *M. smegmatis ΔsigB* based on a cut-off of 2-fold increase or decrease in transcript levels compared to *M. smegmatis* with *sigB* intact (**Figure 1a and 1b**). Of the 585 genes, 149 genes were upregulated and 436 genes were downregulated. In *M. smegmatis* $RbpA_{Mtb}^{R88A}$ there were a total of 233 genes differentially expressed, where 162 of the genes were upregulated and 71 genes were downregulated, indicating that RbpA's interaction with σ^A and σ^B through R88 in most cases represses transcription but can also activate transcription at a smaller subset of genes. Among the total 304 genes upregulated in either of the two strains there was only 7 genes that were upregulated in both strains (**Figure 1a and 1b**). Likewise, among the 491 genes downregulated in either strain, only 16 genes are downregulated in both strains. The minimal overlap in upregulated and downregulated genes in these two strains indicates that during logarithmic growth the σ^B regulon is regulated in a RbpA independent manner.

σ^B regulates small transcripts that if translated would encode short highly charged proteins

To better understand the role of σ^B during logarithmic growth in *M. smegmatis* we analyzed the most downregulated transcripts. Among the 30 most downregulated genes in *M. smegmatis ΔsigB* was a cohort of 20 genes that if translated encode small proteins (**Figure 2**). All 20 of the genes encode for a protein that is less than 100 amino acids in length with an average length of 59 amino acids. Of the 20 genes that would be less than 100 amino acids long, 14 would have an isoelectric point (pI) either below 4.0 or above 8.0. There is an even split between acidic and basic short proteins, with six proteins having a potential pI < 4.0 and eight proteins with a potential pI > 8.0.

Little is known about the function of the σ^B regulated *M. smegmatis* genes that potentially encode small proteins. Most of the gene products do not have known orthologues in other bacterial species and of the 20 small gene products, 16 are annotated as hypothetical or conserved hypothetical in the mycobrowser database. The limited information available for these 16 genes makes it unclear whether these transcripts are translated into short proteins or whether they function as small RNAs. In addition, the lack of orthologous genes suggests that whatever role(s) these gene products play is specific to *M. smegmatis*.

Among the four genes that have annotation information available or a known ortholog in at least one additional mycobacteria species, are two genes, MSMEG_0360 and MSMEG_5635, which are annotated as conserved hypothetical but have orthologs in several other mycobacterial species. MSMEG_0360 is listed as an ortholog of *M. tuberculosis* Rv0236A that has been identified in whole cell lysate and is included in the 'cell wall and cell processes' functional category. A Rv0236A ortholog is found in other mycobacteria species including *M. bovis*, *M. leprae* and *M. marinum*. Though, MSMEG_0360 is likely translated, it is unclear whether

MSMEG_0360 is an essential protein in *M. smegmatis* due to the contradictory predictions of whether Rv0236A is essential in *M. tuberculosis* in two different transposon screens. MSMEG_0360 is not one of the 14 genes that potentially encode an acidic or basic protein, with a neutral pI of 6.6.

M. tuberculosis Rv0909 is orthologous to MSMEG_5635 and orthologs are also found in *M. bovis*, *M. leprae* and *M. marinum*. The protein encoded by this gene has not been detected in whole cell lysate or filtrate and therefore it is unknown whether this gene is translated. Like MSMEG_0360, MSMEG_5635 is potentially a small protein but is not one of the highly charged σ^B regulated genes. MSMEG_5832 is annotated as phosphoribosylformylglycinamidine synthase (which may take the record for longest 'product' annotation on mycobrower…) and has been assigned the gene name *purS*. The *M. tuberculosis*, ortholog Rv0787A is predicted to be essential in *M. tuberculosis* and has been detected in whole cell lysate suggesting that MSMEG_5832 is likely translated in *M. smegmatis*. Like MSMEG_0360 and MSMEG_5635, if the MSMEG_5832 transcript is translated the protein is not highly charged. The annotation for MSMEG_0522 in mycobrowser lists pp24 protein as the gene product and the function of the protein is unknown. MSMEG_0522 does not have known orthologs in other mycobacterial species and if MSMEG_0522 is translated, the protein would be acidic with an isoelectric point of 3.62, however, there is no evidence that this transcript is translated.

sigB is synthetically essential when RbpA activities are limited by point mutations in RbpA BL and SID

Our analysis of the *sigB* regulon during logarithmic growth indicated that RbpA dependent transcription mostly occurs through interactions with the RNAP-σ^A holoenzyme. To compare the roles played by RbpA BL versus RbpA SID during σ^A mediated transcription and

rule out any contribution made by interactions between RbpA and σ^B, we attempted to engineer *M. smegmatis* strains complemented with either *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ and *sigB* deleted. Three different approaches were taken to generate these strains without success (**Figure 3**). In the first approach, we attempted to swap *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ for the *rbpA*$_{Mtb}^{WT}$ gene complemented at the phage attachment site, *attB*, in a strain of *M. smegmatis* lacking both the endogenous *rbpA* and *sigB* loci (**Figure 3a**). For a positive control, we transformed a plasmid carrying *rbpA*$_{Mtb}^{WT}$ into *M. smegmatis* with *rbpA*$_{Mtb}^{WT}$ complemented at the *attB* site (*M. smegmatis rbpA*$_{Mtb}^{WT}$). *M. smegmatis* colonies from these transformations were patched for selection on both zeocin, to select for colonies that likely retained zeocin linked *rbpA*$_{Mtb}^{WT}$ at the *attB* site, and kanamycin to select for colonies that potentially replaced *rbpA*$_{Mtb}^{WT}$ with a kanamycin linked mutant *rbpA* construct or *rbpA*$_{Mtb}^{WT}$ in the case of the positive control. Transformations of *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ into *M. smegmatis rbpA*$_{Mtb}^{WT}$ yielded far fewer colonies compared to the transformations of *rbpA*$_{Mtb}^{WT}$ into either *M. smegmatis ΔsigB rbpA*$_{Mtb}^{WT}$ or *M. smegmatis rbpA*$_{Mtb}^{WT}$ as well as fewer colonies compared to transformations of *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ into *M. smegmatis rbpA*$_{Mtb}^{WT}$ (**Figure 3b**). The few colonies that managed to survive the transformation of plasmids carrying *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ into *M. smegmatis ΔsigB rbpA*$_{Mtb}^{WT}$ and initial selection on kanamycin were then patched for selection on zeocin and kanamycin to confirm the loss of zeocin resistance (Zeo^R). A total of 79 colonies grew from the transformations of *rbpA*$_{Mtb}^{R79A}$ into *M. smegmatis ΔsigB rbpA*$_{Mtb}^{WT}$ and the initial selection on kanamycin. Of these 79 colonies, zero were kanamycin resistant and zeocin sensitive (Kan^R/Zeo^S), and instead all were resistant to both antibiotics (Kan^R/Zeo^R), indicating these colonies had not lost the *rbpA*$_{Mtb}^{WT}$ gene at the *attB* site that occurs with loss of zeocin resistance. Likewise, only 58 colonies grew up from the transformation of *rbpA*$_{Mt}^{R88A}$ into *M.*

smegmatis ΔsigB rbpA$_{Mtb}^{WT}$ and initial kanamycin selection. Patching of the 58 colonies on kanamycin and zeocin yielded 56 Kan^R/Zeo^R colonies and two Kan^R/Zeo^S. To determine whether the two colonies displaying the Kan^R/Zeo^S phenotype had the *rbpA*$_{Mtb}^{WT}$ or the mutant *rbpA*$_{Mtb}^{R88A}$ allele, we sequenced the *attB* site and both Kan^R /Zeo^S colonies had the *rbpA*$_{Mtb}^{WT}$ gene rather than *rbpA*$_{Mtb}^{R88A}$ indicating that through an unknown mechanism these two colonies had managed to acquire Kan^R without the replacement of *rpbA*$_{Mtb}^{WT}$ with *rbpA*$_{Mtb}^{R88A}$. The inability to swap *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ for *rbpA*$_{Mtb}^{WT}$ in *M. smegmatis ΔsigB rbpA*$_{Mtb}^{WT}$ stands in sharp contrast to the efficiency of swapping *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ for *rpbA*$_{Mtb}^{WT}$ in *M. smegmatis rbpA*$_{Mtb}^{WT}$ as well as the efficiency of swapping *rbpA*$_{Mtb}^{WT}$ for *rbpA*$_{Mtb}^{WT}$ in *M. smegmatis ΔsigB rbpA*$_{Mtb}^{WT}$.

The second approach to engineer *M. smegmatis ΔsigB rbpA*$_{Mtb}^{R79A}$ or *M. smegmatis ΔsigB rbpA*$_{Mtb}^{R88A}$ involved deleting the endogenous *rbpA* locus in a *M. smegmatis* strains lacking *sigB* and already complemented with either *rbpA*$_{Mtb}^{R79A}$ or *rbpA*$_{Mtb}^{R88A}$ at the *attB* site (**Figure 3c**). For this approach, we attempted to delete the endogenous *rbpA* locus using two step allelic exchange. This process includes an initial transformation of a plasmid containing a positive selection marker that confers *M. smegmatis* resistance to hygromycin (Hyg^R) as well as two negative selection markers that confer *M. smegmatis* sensitivity to sucrose and 2-deoxygalactose (2-DOG). The transformants are first selected for hygromycin resistance indicating that the plasmid has recombined into the bacterial chromosome and these colonies are referred to as intermediates (**Figure 3d**). The plasmid that has recombined into the intermediate's chromosome has sites of homology flanking *rbpA* that can recombine allowing the plasmid to remove itself from the chromosome. Sites of homology within the plasmid and bacterial chromosome are located such that recombination for plasmid removal can result in *rbpA* deletion or restoration of

the bacterial chromosome to the status prior to integration with equal probability for either outcome. Only the intermediates which underwent a second recombination event that removes the plasmid survive negative selection on agar containing sucrose and 2-DOG. The colonies capable of growing in the presence of sucrose and 2-DOG are then patched on LB-agar containing hygromycin and LB-agar containing both sucrose and 2-DOG to confirm the loss of Hyg^R. The Hyg^S /Suc 2-DOG^R patches are *rbpA* deletion candidates. We determine whether the deletion candidates reverted to their original chromosome composition or lost *rbpA* during the second recombination event by southern blotting.

The data collected from this second approach is limited but so far supports the hypothesis that *sigB* is made synthetically essential by the RbpA R79A and R88A substitutions. A technical hurdle of this approach is the low efficiency of the plasmid DNA recombination into the bacterial chromosome. So far only a total of five intermediates have grown from transformations into either *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{R79A}$, *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{R88A}$ or *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{WT}$ and selection for Hyg^R (**Figure 3d**). However, four of the intermediates originated from the transformation into *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{R79A}$, which has yielded at least 26 deletion candidates. Another challenge of this approach is the lower throughput screening of the deletion candidates with southern blotting and therefore only six candidates have been screened so far. All six of the candidates had an intact *rbpA* endogenous locus. Although only six deletion candidates have been screened and more screening will be required to strengthen the conclusions drawn from this approach, the equal probability of the second recombination event yielding the original chromosome and deletion of *rbpA* means that we would have expected three of the six candidates to have lost *rbpA* during the second recombination event. Statistically, assuming that the outcome of the second recombination event

is binary and both possibilities are equally likely to occur then the probability that we would observe the same outcome for all six candidates is less than 2%. Moreover, among the six candidates screened for deletion of *rbpA* in *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{WT}$, *rbpA* was deleted in two of the candidates, which is much closer to the expected 50% deletion rate that should occur when a gene is non-essential.

The third approach involved deleting *sigB* from *M. smegmatis* strains complemented with mutant *rbpA* constructs at the *attB* site and the endogenous *rbpA* locus deleted (**Figure 3e**). This approach utilizes a mycobacteria phage recombineered to target the *M. smegmatis sigB* locus and successful deletion of *sigB* occurs through a linear combination event that replaces the *sigB* locus with a hygromycin resistance cassette. Of the three approaches, I have collected the least amount of data for this approach and so far the data does not lend support either for or against the hypothesis that *sigB* is made synthetically essential by the RbpA substitutions R79A and R88A. Similar to the second approach, the screening of deletion candidates, in this case for the deletion of *sigB*, is done through low-throughput southern blotting. Only *sigB* deletion candidates that have come from the phage infection of the control *M. smegmatis attB::rbpA*$_{Mtb}^{WT}$ have been screened. Zero out of six Hyg^{R} *M. smegmatis attB::rbpA*$_{Mtb}^{WT}$ *sigB* deletion candidates had a *sigB* deletion, probably indicating a problem with the phage recombineering and the need to troubleshoot the phage production.

Discussion

The findings of this study shed new light on the role of *sigB* in *M. smegmatis*. We determine the *sigB* regulon and find that *sigB* regulates almost 1/10th of the *M. smegmatis* transcriptome during logarithmic growth in a mostly RbpA independent manner (**Figure 1**). We find that among the 30 genes most downregulated in the *M. smegmatis sigB* deletion strain are

genes that potentially encode functionally uncharacterized small proteins (**Figure 2**). Most of these genes do not have known orthologues in other bacterial species and if translated most of these proteins have extreme isoelectric points. In addition to determining that σ^B functions in a RbpA independent manner during logarithmic growth our data suggests that *sigB* is made synthetically essential by loss of RbpA activities that require RbpA BL and SID (Figure 3).

This data adds to the notion that the function of *sigB* in mycobacteria is distinct from that of *rpoS* in *E. coli*. In *E. coli*, σ^S and σ^{70} have distinct roles with σ^{70} controlling transcription during conditions permissive to fast growth and division while σ^S is thought to take control of transcription for survival in challenging environments. In mycobacteria, *sigB* is not only at levels comparable to *sigA* during logarithmic growth but now our data indicates σ^B serves an active role during logarithmic growth, contributing to the transcription of almost 600 genes. Our observation that σ^B regulates many *M. smegmatis* genes during logarithmic growth aligns with the combined RNA-seq and ChIP-seq analysis of the *M. tuberculosis* σ^B regulon showing that σ^B regulates housekeeping genes during logarithmic growth [143]. Another potential hint that σ^B is involved in housekeeping transcription comes from our finding that σ^B is synthetically essential when RbpA activities are limited. The reason why σ^B is conditionally essential is not yet clear but one possible explanation is σ^B's contribution to housekeeping transcription becomes essential when limited RbpA activity results in diminished transcription of housekeeping genes. Collectively, this data suggests that mycobacteria σ^B is not only involved in responding to stress similar to the *E. coli* σ^S paradigm but is also contributing to fast growing mycobacteria transcription in a nutrient rich stress-free environment.

Previously we showed that the R88A substitution weakens RbpA binding to both σ^A and σ^B, which would suggest that whatever the σ^B's essential function is in the context of the RbpA

R88A mutation is RbpA independent [125]. However, recent work characterizing the impact of substitutions targeting charged residues in RbpA BL found that the R79A substitution did not affect RbpA enhancement of σ^B activity, whereas other substitutions in RbpA including K73A and K74A weaken RbpA enhancement of σ^B activities [148]. These results suggest that RbpA BL alters σ^A and σ^B activities through distinct mechanisms. In this case, perhaps the σ^B-RbpA R79A complex is still capable of carrying out functions that σ^A-RbpA R79A cannot. If σ^B-RbpA R79A is now responsible for an essential activity normally carried out by σ^A-RbpA then this could explain why deletion of σ^B is lethal in the R79A genetic context.

Analysis of the 30 genes most downregulated in the *M. smegmatis sigB* deletion strain revealed that most of this cohort consists of small gene products encoding either short untranslated RNAs or small proteins with an average length of 59 amino acids. If these genes are translated, 16 of the 20 small proteins would be highly charged. For the most part, little is known about these 20 genes and the lack of known orthologues in other bacterial species suggests that they are unique to *M. smegmatis*.

Molecular dissection of the *E. coli* σ^S regulatory network has shown that small regulatory RNAs are at the center of controlling σ^S levels and fine-tuning the general stress response to address a specific stress [136]. In many cases, the small regulatory RNAs bind to the *sigS* transcript resulting in either stabilization or degradation of the *sigS* transcript. To my knowledge, it is unclear whether mycobacteria *sigB* is regulated post-transcriptionally through small RNAs. Future work that could provide insight into this question is determining whether the σ^B regulated small gene transcripts have stretches of complementary sequences that would base pair with the *sigB* transcript indicating that these small RNAs could bind and potentially regulate *sigB* post-transcriptionally. In addition to binding the *sigB* transcript, the small genes could regulate *sigB* or

σ^B by binding to the transcripts of genes responsible for degrading the *sigB* transcript or σ^B proteolysis.

If future work determines that the σ^B regulated small genes are translated, then this opens the door to another set of interesting potential functions. Small highly charged proteins have been identified in many other bacterial species and have been found to serve numerous functions including but not limited to bacterial spore formation, stress signaling, efflux, cell division, quorum sensing, bacteriocin production that is utilized for bacterial chemical warfare and antibiotic resistance [151]. The list of different types of small proteins and the functions they perform continues to expand providing many starting points for future work defining the function of *M. smegmatis* σ^B regulated small proteins.

Figures

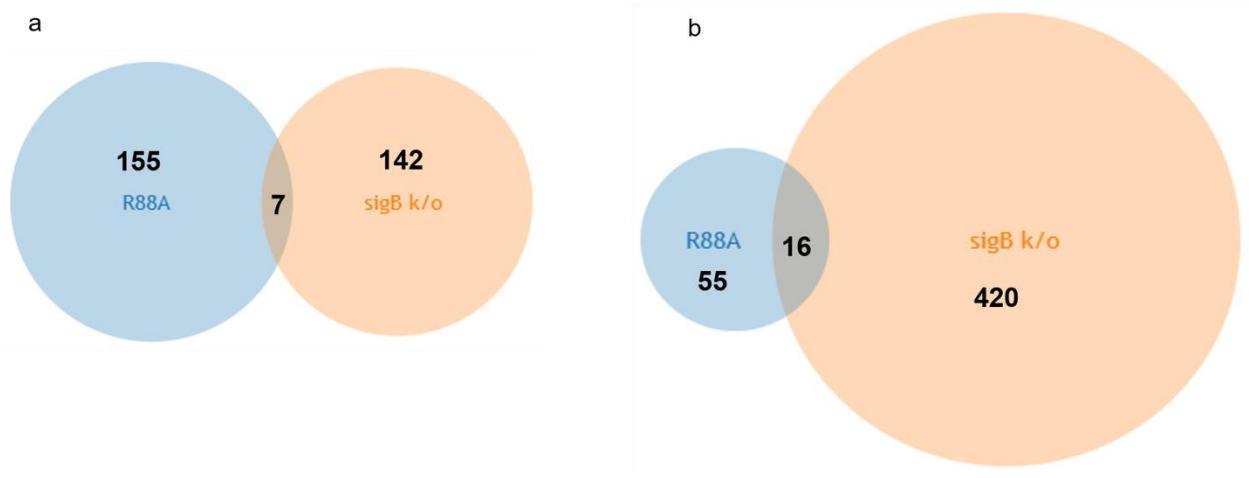


Figure 1. Overlap comparison of the upregulated and downregulated genes in *M. smegmatis ΔsigB* and *M. smegmatis* $RbpA_{Mtb}^{R88A}$.

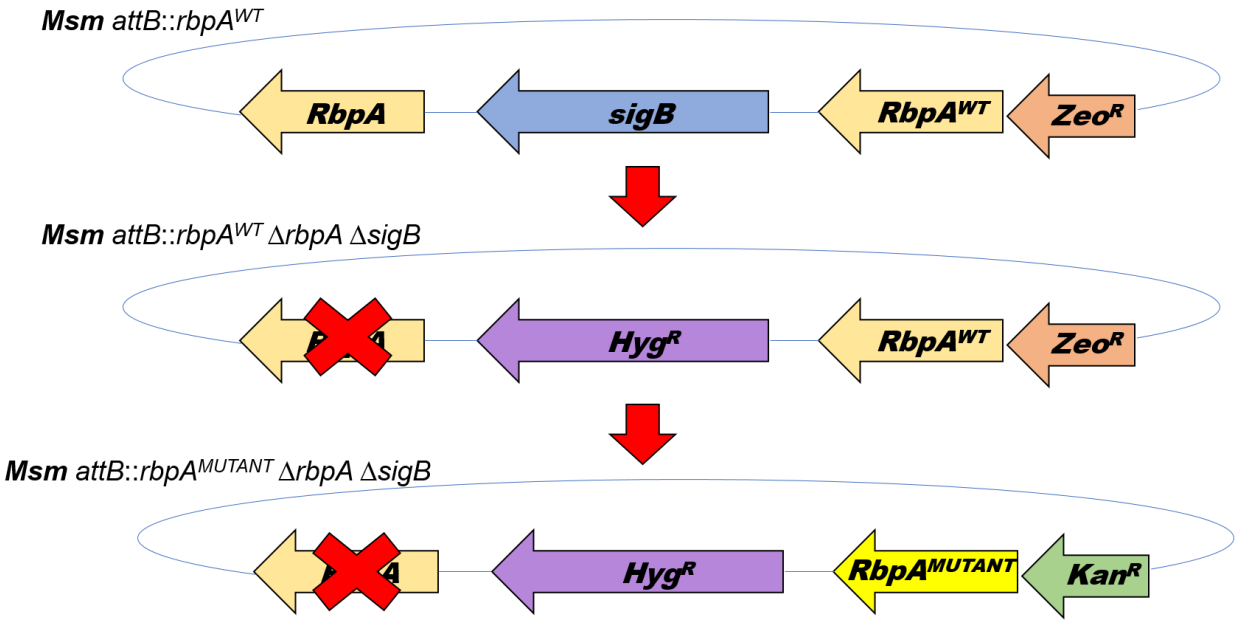
a. Venn diagram showing the number of the genes that are upregulated at least 2-fold in both *M. smegmatis ΔsigB* and *M. smegmatis* $rbpA_{Mtb}^{WT}$.

b. Venn diagram showing the number of the genes that are downregulated at least 2-fold in both *M. smegmatis ΔsigB* and *M. smegmatis* $rbpA_{Mtb}^{WT}$.

#	*Msm* ID	Fold Change	Avg. Read Count	Size (aa)	pI	#	*Msm* ID	Fold Change	Avg. Read Count	Size (aa)	pI
1	MSMEG_2752	-6.26651	10442.4	319		16	MSMEG_5527	-2.5952	769.0	227	
2	MSMEG_6211	-3.38498	1617.0	50	4.19	17	MSMEG_1758	-2.38817	413.7	57	3.8
3	MSMEG_5180	-3.06247	563.1	43	3.60	18	MSMEG_5875	-2.1446	366.2	56	4.0
4	MSMEG_0879	-2.62127	813.7	48	11.5	19	MSMEG_5558	-2.29581	1018.1	103	
5	MSMEG_5832	-2.40976	1188.8	82	4.58	20	MSMEG_5343	-1.61222	12988.1	115	
6	MSMEG_0945	-1.86703	1549.0	33	12.9	21	MSMEG_1802	-1.35976	10630.1	139	
7	MSMEG_1788	-2.17058	5707.1	84	8.77	22	MSMEG_1599	-1.45912	2172.3	192	
8	MSMEG_5195	-3.48403	136.3	47	12.0	23	MSMEG_0651	-3.51421	309.8	226	
9	MSMEG_0522	-3.5613	227.2	50	3.62	24	MSMEG_5169	-2.57383	242.6	56	11.1
10	MSMEG_1783	-2.23196	1749.4	76	3.50	25	MSMEG_5969	-1.92095	1825.2	404	
11	MSMEG_0360	-2.33359	2348.8	57	6.6	26	MSMEG_0009	-3.44889	109.5	30	12.2
12	MSMEG_1770	-2.4379	55624.0	80	5.3	27	MSMEG_4693	-1.51468	1646.6	79	7.3
13	MSMEG_5355	-1.70676	1468.1	191		28	MSMEG_4961	-1.40262	2224.1	91	10.6
14	MSMEG_3443	-2.70814	725.4	52	3.8	29	MSMEG_5298	-2.39116	442.1	104	
15	MSMEG_4224	-3.46159	157.6	42	12.5	30	MSMEG_5635	-2.42444	1234.7	62	4.7

Figure 2. List of the 30 most downregulated genes in *M. smegmatis ΔsigB*. Genes highlighted in yellow would be less than 100 amino acids long if translated.

a

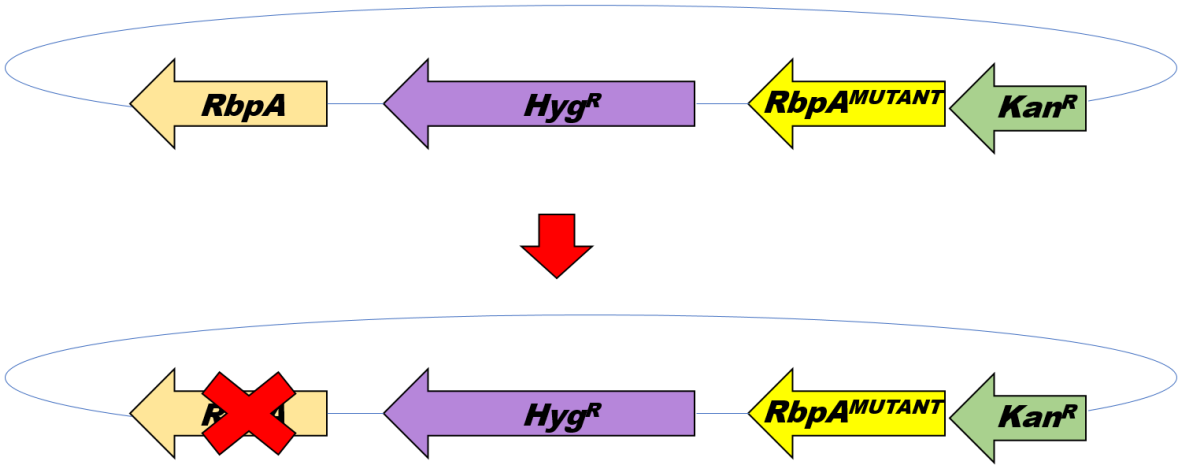


b

Strain transformed	RbpA construct transformed	Number of cells transformed	Number of patches	Kan^R/Zeo^S
csm291 (Δ*rbpA*)	*wt rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	100	99
csm413 (Δ*rbpA* and Δ*sigB*)	*wt rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	100	100
csm291 (Δ*rbpA*)	*R79A rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	100	100
csm413 (Δ*rbpA* and Δ*sigB*)	*R79A rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	79	0
csm291 (Δ*rbpA*)	*R88A rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	100	93
csm413 (Δ*rbpA* and Δ*sigB*)	*R88A rbpA*	$8 \times 10^8 – 1.2 \times 10^9$	58	*2

*I have sequenced the *attB* site of the 2 csm413 R88A transformants and *wt rbpA* is present.

c



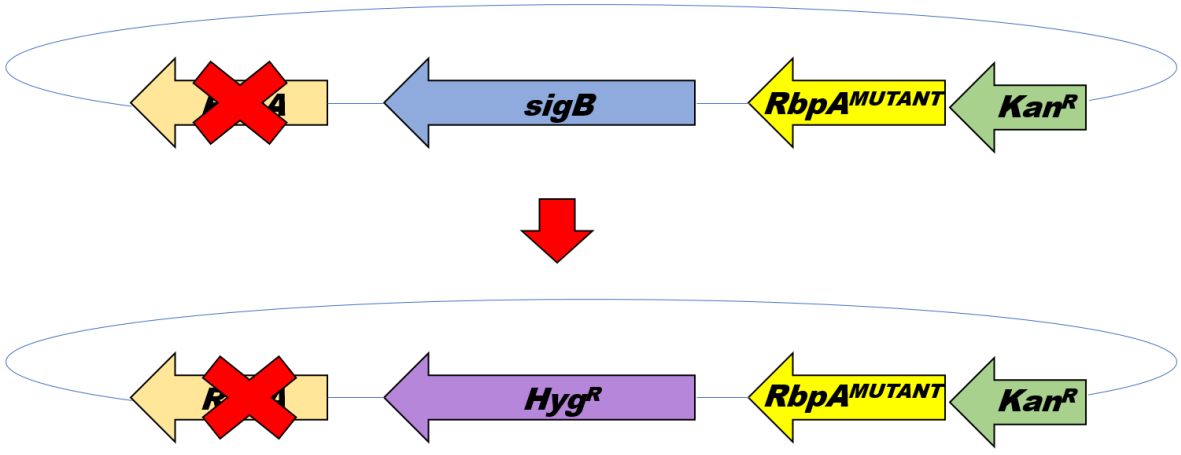
d

Strain transformed (RbpA/CarD construct)	Number of cells transformed	Number of colonies (*intermediates)	Number of *deletion candidates	Δ*rbpA* of deletion candidates screened
csm419 (RbpA WT)	1.51 x 10^9	1	12	2/6
csm418 (RbpA R88A)	1.18 x 10^9	0	0	0
csm417 (RbpA R79A)	1.22 x 10^9	4	26	0/6
csm452 (RbpA R27E)	1.08 x 10^9	2	33	0/6
csm453 (RbpA 26-111)	1.07 x 10^9	0	0	0

*Intermediates – colonies that have undergone the first recombination

*Deletion candidates – colonies that have undergone both recombinations, confirmed by selection for Suc2DOG^R/Hyg^S

e



f

Strain infected (RbpA/CarD construct)	Number of cells infected	Number of colonies	Δ*sigB* of colonies screened
csm275 (RbpA WT)	3.43×10^9	6	0/6
csm314 (RbpA R88A)	2.88×10^9	0	0
csm322 (RbpA R79A)	2.4×10^9	0	0
csm448 (RbpA R27E)	3.1×10^9	0	0
mgm3044 (CarD R25E)	3.65×10^9	2	0/2

Figure 3. Genetic approaches used in the attempt to engineer *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{R79A}$ and *M. smegmatis ΔsigB attB::rbpA*$_{Mtb}^{R88A}$

Chapter 5: Conclusions

Jerome Prusa

Major Findings

The activities of all four RbpA structural domains are required for *M. tuberculosis* viability while only the RbpA BL and SID activities are sufficient for *M. smegmatis* viability but required for normal *M. smegmatis* growth

To determine which of the four RbpA structural domains are essential for *M. tuberculosis* growth we engineered a cohort of RbpA mutants targeting each domain and tested whether the mutant RbpA constructs can support viability in either *M. tuberculosis* or *M. smegmatis*. The two RbpA truncation mutant constructs included in this analysis were $RbpA_{Mtb}^{1-71}$, which lacks both the C-terminal basic linker (BL) and sigma interaction domain (SID), and $RbpA_{Mtb}^{72-111}$, which is the inverse of $RbpA_{Mtb}^{1-71}$ and lacks the N-terminal tail (NTT) and core domain (CD). We found that neither RbpA truncation mutant supported *M. tuberculosis* viability while only $RbpA_{Mtb}^{72-111}$ supported *M. smegmatis* viability suggesting that all four RbpA structural domains are required for *M. tuberculosis* viability while only the BL and SID domains are required for *M. smegmatis* viability **(Chapter 2 Figure 1a)**[125]. Prior work had shown that the R79 residue located in the RbpA BL is involved in RbpA's interaction with DNA [22], while R88 located in RbpA's SID is a key residue for *M. tuberculosis* RbpA binding to σ^A and σ^B [22]. To determine whether RbpA BL and SID are essential in *M. smegmatis* and *M. tuberculosis* due to the activities mediated by R79 and R88, we engineered the point mutant constructs $RbpA_{Mtb}^{R79A}$ and $RbpA_{Mtb}^{R88A}$ and tested whether $RbpA_{Mtb}^{R79A}$ and $RbpA_{Mtb}^{R88A}$ are viable in *M. smegmatis* and *M. tuberculosis*. Both RbpA point mutations are lethal in *M. tuberculosis* **(Chapter 2 Figure 1a)** [125]. This result shows that RbpA activities mediated by R79 and R88 are at least one reason why the BL and SID are essential in *M. tuberculosis*. In *M. smegmatis*, both RbpA point mutations are viable but confer a

slow growth phenotype showing that R79 and R88 dependent RbpA activities are required for normal *M. smegmatis* growth **(Chapter 2 Figure 1b and 1c)**[125].

RbpA R88 is necessary and sufficient for RbpA binding to mycobacteria RNAP-σ^A and RNAP-σ^B holoenzyme

RbpA makes several interactions with the RNAP-σ^A holoenzyme and is predicted to interact with the RNAP-σ^B holoenzyme in a similar fashion [22,67,106]. To determine which RbpA structural domains are required for RbpA binding to the RNAP-σ^A and RNAP-σ^B holoenzymes we engineered *M. smegmatis* strains expressing FLAG tagged versions of RbpA$_{Mtb}^{WT}$, RbpA$_{Mtb}^{R79A}$, RbpA$_{Mtb}^{R88A}$ and RbpA$_{Mtb}^{72\text{-}111}$ and compared the levels of RNAP β subunit, as a read out for the core RNAP, σ^A and σ^B co-immunoprecipitated (co-IPed) by each RbpA construct. RbpA$_{Mtb}^{R79A}$-FLAG co-IPed similar levels of RNAP β, σ^A and σ^B indicating that the interactions mediated by RbpA R79 with either the RNAP-σ^A or RNAP-σ^B holoenzyme are not required for RbpA binding to either holoenzyme **(Chapter 2 Figure 2a – 2c)**[125]. RbpA$_{Mtb}^{R88A}$-FLAG co-IPed almost undetectable levels of σ^A and σ^B, in agreement with previous data showing that the R88 equivalent in the *S. coelicolor* RbpA orthologue is required for RbpA binding to *S. coelicolor*'s principal housekeeping σ factor and principal-like alternative σ factor **(Chapter 2 Figure 2a – 2c)**[110,125]. In addition to σ^A and σ^B, RbpA$_{Mtb}^{R88A}$-FLAG also co-IPed low levels of RNAP β that are comparable to the background amount RNAP β co-IPed by untagged RbpA$_{Mtb}^{WT}$ **(Chapter 2 Figure 2a and 2d)**[125]. RbpA R88 is not positioned to interact with RNAP β or any of the other core RNAP subunits and is only positioned to interact with σ^A and σ^B [22,67,106]. Based on the predicted lack of RbpA R88 interaction with other RNAP subunits and our finding that RbpA$_{Mtb}^{R88A}$ co-IPed lower amounts of all the RNAP subunits we conclude that the RbpA R88 dependent interaction with σ^A or σ^B is necessary and sufficient for RbpA's

interaction with the RNAP-σ^A and RNAP-σ^B holoenzymes. This assertion is strengthened by our additional finding that $RbpA_{Mtb}^{72\text{-}111}$-FLAG co-IPs the same level of RNAP β, σ^A and σ^B as $RbpA_{Mtb}^{WT}$-FLAG, showing that neither RbpA NTT or CD are required for RbpA binding to RNAP-σ^A or RNAP-σ^B holoenzymes **(Chapter 2 Figure 2a – 2d)**[125].

RbpA both promotes and antagonizes RNAP open promoter complex stability in a domain dependent manner

An *in vitro* comparison of RNAP open promoter complex (RP_o) stability between *E. coli* RNAP-σ^{70} holoenzyme and *M. bovis* RNAP-σ^A holoenzyme on the *M. tuberculosis* ribosomal rRNA AP3 (*rrn*AP3) promoter showed that *E. coli* RNAP-σ^{70} holoenzyme RP_o stability is much higher compared to its *M. bovis* counterpart [115]. Addition of *M. tuberculosis* essential RNAP binding protein CarD to the *M. bovis* RNAP-σ^A holoenzyme increased RP_o stability. Likewise, RbpA increases RP_o stability and together both CarD and RbpA increase *M. bovis* RNAP-σ^A RP_o stability to a level higher than either factor individually and to a level that is comparable to *E. coli* RNAP-σ^{70} RP_o [65]. To determine which RbpA domains increase RP_o stability, we compared *M. tuberculosis* RNAP-σ^A RP_o stability on the *M. tuberculosis rrn*AP3 promoter when bound by either $RbpA_{Mtb}^{WT}$, $RbpA_{Mtb}^{R79A}$, $RbpA_{Mtb}^{R88A}$ or $RbpA_{Mtb}^{72\text{-}111}$. RbpA RP_o stabilizing activity requires both R79 and R88, as a substitution at either residue results in complete loss of RbpA RP_o stabilizing activity **(Chapter 2 Figure 3a and 3b)**[125]. Conversely, the truncation of RbpA NTT and CD increased RbpA RP_o stabilizing activity compared to $RbpA_{Mtb}^{WT}$ showing that either RbpA NTT or CD or both domains antagonize *M. tuberculosis* RNAP-σ^A RP_o stability. To further dissect which RbpA domain antagonizes RP_o stability, we engineered $RbpA_{Mtb}^{26\text{-}111}$, which only lacks the RbpA NTT and measured *M. tuberculosis* RNAP-σ^A holoenzyme RP_o stability when bound by $RbpA_{Mtb}^{26\text{-}111}$. The RP_o stabilizing activity of $RbpA_{Mtb}^{26\text{-}111}$ was greater

than RbpA$_{Mtb}^{WT}$ but less than RbpA$_{Mtb}^{72\text{-}111}$ showing that both RbpA NTT and CD antagonize *M. tuberculosis* RNAP-σ^A RP$_o$ stabilizing activity **(Chapter 2 Figure 3a and 3b**)[125].

RbpA BL and SID activities can repress and activate mycobacteria transcription in a gene dependent manner

To determine which *M. smegmatis* genes are regulated by RbpA BL and SID dependent activities we compared the transcriptomes of *M. smegmatis* strains expressing either RbpA$_{Mtb}^{WT}$, RbpA$_{Mtb}^{R79A}$ or RbpA$_{Mtb}^{R88A}$. RNA was collected from the *M. smegmatis* strains grown to mid-logarithmic phase and submitted for RNA-sequencing (RNA-seq). Both upregulated and downregulated genes were observed in both the RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ strains indicating that R79 and R88 dependent RbpA BL and SID activities in some cases activate transcription and in other cases repress transcription **(Chapter 2 Figure 4b)**[125]. Of the 103 genes significantly upregulated at least 2-fold in the *M. smegmatis* strain expressing RbpA$_{Mtb}^{R79A}$, 81 of these genes were also downregulated at least 2-fold in the *M. smegmatis* strain expressing RbpA$_{Mtb}^{R88A}$. Likewise, of the 96 genes downregulated at least 2-fold in the RbpA$_{Mtb}^{R79A}$ expressing *M. smegmatis*, 34 of those genes were also downregulated at least 2-fold in RbpA$_{Mtb}^{R88A}$. The large degree of overlap in upregulated and downregulated genes between these strains shows that disrupted RbpA BL and SID activities results in a similar effect on transcription at various promoters across the *M. smegmatis* genome **(Chapter 2 Figure 4c and 4d**)[125].

Truncation of RbpA NTT and CD reduces RNA levels in *M. smegmatis*

The RNA-seq results showed that a much larger cohort of genes is differentially expressed as a result of RbpA NTT and CD truncation compared to loss of R79 and R88 dependent RbpA activities from *M. smegmatis* **(Chapter 2 Figure 4b)**[125]. The RNA-seq results also showed that the genes that are differentially expressed in *M. smegmatis* RbpA$_{Mtb}^{72\text{-}111}$ are

different from those that are differentially expressed in either *M. smegmatis* expressing $RbpA_{Mtb}^{R79A}$ or *M. smegmatis* expressing $RbpA_{Mtb}^{R88A}$. However, similar to the effect of RbpA BL or SID point mutations, the truncation of RbpA NTT and CD results in activation of some genes and repression of other genes **(Chapter 2 Figure 4b)**[125]. After completing the RNA-seq comparison, we recognized that one potentially important caveat of the RNA-seq experimental design is that equivalent µg amounts of RNA from each biological replicate were submitted for sequencing without considering how much total RNA was isolated from each biological replicate. Under this experiment design, a mutation or condition that increases or decreases the overall amount of RNA per bacterium would be masked. For example, if mutation A reduces the total amount of RNA per bacterium by one half compared to mutation B, we would submit RNA from 2X the number of bacterium with mutation A as that from mutation B in order to provide an equivalent µg amount of RNA for sequencing from each replicate. Most RNA-seq experiments compare changes in a transcriptome resulting from mutations or conditions that are not known to alter a cell's transcriptional machinery and therefore this caveat is often not considered. However, because we know that RbpA is a part of the housekeeping transcription machinery in *Actinobacteria* we tested whether the point mutations in RbpA BL or RbpA SID or the truncation of both RbpA NTT and CD results in an overall increase or decrease in RNA per bacterium. We did this by 'spiking-in' a known µg amount of MS2 phage RNA into the total RNA collected where the amount of spike-in RNA added is determined by the number of bacteria that the RNA was collected from.

Throughout the process of collecting RNA for RNA-sequencing I noticed that I was repeatedly extracting less RNA from *M. smegmatis* $RbpA_{Mtb}^{72\text{-}111}$ samples compared to samples from the other strains despite knowing that I was collecting RNA from an equivalent number of

bacteria based on the culture volume and OD_{600}. Due to this observation, our spike-in analysis was guided by the suspicion that the amount of RNA per bacterium is lower in *M. smegmatis* $RbpA_{Mtb}^{72\text{-}111}$. When we used qRT-PCR and normalization to the MS2 spike-in control to measure the transcript levels of eight genes that are not differentially expressed in *M. smegmatis* $RbpA_{Mtb}^{72\text{-}111}$ according to the RNA-seq, we found that all eight genes were downregulated compared to *M. smegmatis* $RbpA_{Mtb}^{WT}$ **(Chapter 2 Figure 4e)**[125]. To further investigate this discrepancy between the RNA-seq and qRT-PCR spike-in control results, we then analyzed the transcript levels of the eight most upregulated genes in *M. smegmatis* $RbpA_{Mtb}^{72\text{-}111}$ according to the RNA-seq and found that these genes are also downregulated compared to *M. smegmatis* $RbpA_{Mtb}^{WT}$. The downregulation of all 16 genes evaluated by qRT-PCR, provides strong evidence in support of our hypothesis that the truncation of RbpA NTT and CD results in overall decrease in transcription

When we measured the transcript levels of these same 16 genes in *M. smegmatis* $RbpA_{Mtb}^{R79A}$ and *M. smegmatis* $RbpA_{Mtb}^{R88A}$ using the qRT-PCR spike in control method, the results from the RNA-seq and qRT-PCR spike in control agreed **(Chapter 2 Figure 4e)**[125]. In this case, the agreement between the two techniques provided confirmation of the RNA-seq results and supported the conclusion that the loss of R79 and R88 dependent RbpA BL and SID activities does not reduce overall levels of RNA per bacterium.

RbpA R10 and RbpA E17 synergize to promote fidaxomcin inhibition of *M. tuberculosis* RNAP-σ^A activity *in vitro* while other conserved RbpA NTT residues including R4, L6, R7 and S15, RbpA CD, RbpA R79 and RbpA R88 do not

Fidaxomicin (Fdx) inhibits *M. tuberculosis* RNAP-σ^A activity *in vitro* and *M. smegmatis* growth in a RbpA NTT dependent manner [67]. The structure of Fdx bound to *M. tuberculosis*

RNAP-σ^A in complex with RbpA shows that RbpA NTT is positioned near the Fdx binding pocket and RbpA NTT residues R10 and E17 interact with Fdx [67]. Based on these data, Boyaci et al. hypothesized that RbpA NTT dependent Fdx activity is due to RbpA NTT's contribution to the binding interface between Fdx and *M. tuberculosis* RNAP-σ^A. To further interrogate the mechanism of RbpA NTT dependent Fdx activity and test whether RbpA R10 and E17 increase Fdx activity we measured the *in vitro* Fdx activity against *M. tuberculosis* RNAP-σ^A bound by RbpA mutant constructs including $RbpA_{Mtb}^{R10A}$, $RbpA_{Mtb}^{E17A}$ or $RbpA_{Mtb}^{R10A/E17A}$. Individually, the alanine substitution of RbpA R10 and E17 did not change Fdx activity against the *M. tuberculosis* RNAP-σ^A holoenzyme when compared to Fdx activity against $RbpA_{Mtb}^{WT}$ bound *M. tuberculosis* RNAP-σ^A **(Chapter 3 Figure 1a and 1b)**. However, alanine substitutions of both R10 and E17 occurring in the double point mutant $RbpA_{Mtb}^{R10A/E17A}$ construct, decreased Fdx activity against *M. tuberculosis* RNAP-σ^A by approximately three-fold. These findings show that RbpA NTT dependent activity against the *M. tuberculosis* RNAP-σ^A is partially due to combination of RbpA R10 and E17 but that additional RbpA NTT residues are involved in Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity.

To identify other RbpA NTT residues that play a role in RbpA NTT dependent Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity we engineered a cohort of mutant RbpA constructs targeting four of the most highly conserved RbpA NTT residues, including R4, L6, R7 and S15 [106]. When we measured Fdx inhibition of the *M. tuberculosis* RNAP-σ^A holoenzyme in the presence of either $RbpA_{Mtb}^{R4A}$, $RbpA_{Mtb}^{R4E}$, $RbpA_{Mtb}^{L6A}$, $RbpA_{Mtb}^{R7A}$, $RbpA_{Mtb}^{R7E}$ or $RbpA_{Mtb}^{S15A}$ we observed the same level of Fdx inhibition of $RbpA_{Mtb}^{WT}$ bound *M. tuberculosis* RNAP-σ^A activity, suggesting that R4, L6, R7 and S15 are not individually required for RbpA NTT dependent Fdx activity **(Chapter 3 Figure 3c and 3d)**.

To investigate whether the activities of RbpA structural domains besides the NTT affect Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity, I measured Fdx activity against $RbpA_{Mtb}^{72-111}$, $RbpA_{Mtb}^{R79A}$ or $RbpA_{Mtb}^{R88A}$ bound *M. tuberculosis* RNAP-σ^A. In the presence of $RbpA_{Mtb}^{R79A}$ or $RbpA_{Mtb}^{R88A}$, Fdx inhibition of the *M. tuberculosis* RNAP-σ^A holoenzyme was the same as in the presence of $RbpA_{Mtb}^{WT}$ showing that R79 dependent BL activities, R88 dependent SID activities do not contribute to Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity. $RbpA_{Mtb}^{72-111}$ and $RbpA_{Mtb}^{26-111}$ permit similar levels of Fdx inhibition indicating that the CD does not affect Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity *in vitro* **(Chapter 3 Figure 3e)**.

Multiple RbpA domains and CarD impact fidaxomicin inhibition of *M. smegmatis* growth

Our *in vitro* analysis of Fdx inhibition of *M. tuberculosis* RNAP-σ^A activity showed that only the truncation of RbpA NTT or alanine substitution of both R10 and E17 decrease Fdx activity **(Chapter 3 Figure 1a and 1b)**. Based on these results we predicted that Fdx would be equally lethal against *M. smegmatis* expressing $RbpA_{Mtb}^{WT}$ and the rest of the *M. smegmatis* strains expressing the mutant RbpA constructs we interrogated *in vitro,* with the exception of *M. smegmatis* strains expressing either $RbpA_{Mtb}^{26-111}$, $RbpA_{Mtb}^{72-111}$ or $RbpA_{Mtb}^{R10A/E17A}$ that we expected to be less sensitive to Fdx. Contrary to our expectations, several *M. smegmatis* strains expressing various mutant RbpA constructs showed either an increase or decrease in sensitivity to Fdx compared to *M. smegmatis* $RbpA_{Mtb}^{WT}$ **(Chapter 3 Figure 3)**. Among the strains expressing RbpA NTT point mutants, $RbpA_{Mtb}^{R4A}$ and $RbpA_{Mtb}^{R7A}$ were more sensitive to Fdx at each concentration included in the analysis. At lower Fdx concentrations, *M. smegmatis* $RbpA_{Mtb}^{R4E}$ was less sensitive to Fdx. Also contrary to our predictions based on the *in vitro* results was the finding that *M. smegmatis* $RbpA_{Mtb}^{R10A/E17A}$ and *M. smegmatis* $RbpA_{Mtb}^{WT}$ are

equally sensitive to Fdx. Strikingly, *M. smegmatis* RbpA$_{Mtb}^{R79A}$ and RbpA$_{Mtb}^{R88A}$ were significantly more sensitive to Fdx treatment. These *in vivo* results show that RbpA impacts Fdx inhibition of *M. smegmatis* growth in a way that is not recapitulated in the *in vitro* assay.

The observation that RbpA residues which are not predicted to directly interact with Fdx can alter *M. smegmatis* sensitivity to Fdx led us to investigate whether CarD, which is also not predicted to interact with Fdx, affects *M. smegmatis* sensitivity to Fdx. A glutamate substitution at CarD R25 in CarD's RNAP interaction domain weakens CarD binding to the RNAP β subunit and results in dysregulated transcription at over half of *M. tuberculosis*'s genes [94,97,99]. We compared *M. smegmatis* CarD$_{Mtb}^{R25E}$ and *M. smegmatis* CarD$_{Mtb}^{WT}$ sensitivity to Fdx and found that loss of R25 dependent CarD activity significantly increases *M. smegmatis* sensitivity to Fdx **(Chapter 3 Figure 3d and 3e)**. This result shows that *M. smegmatis* sensitivity to Fdx occurs through both RbpA dependent and independent mechanisms.

RbpA NTT residues positioned near the RNAP β' lid, RNAP σ^A region 3.2, and RNAP β' ZBD antagonize RbpA RP_o stabilizing activity

We showed that RbpA NTT antagonizes RbpA RP_o stabilizing activity through an unknown mechanism **(Chapter 2 Figure 3)**[125]. To gain insight into this mechanism, we investigated how RbpA NTT residues positioned near different regions of the RNAP contribute to RbpA NTT antagonism of RNAP-σ^A RP_o stability. We determine that alanine substitution of L6, positioned near hydrophobic residues in the RNAP β' lid, RNAP β Sw3 and RNAP σ^A region 3.2 and R7 increase RbpA RP_o stabilizing activity **(Chapter 3 Figure 4a and 4b)**. In addition, alanine substitutions at RbpA S15 and E17 positioned near the RNAP β' zinc binding domain also increase RbpA RP_o stabilizing activity. These results suggest that RbpA NTT

antagonism of RP_o stability could occur through RbpA NTT interactions with RNAP β' lid, RNAP β Sw3, RNAP σ^A region 3.2, and RNAP β' ZBD.

RbpA NTT promotes full-length transcription

It is unknown why the amount of RNA per cell is decreased in *M. smegmatis* $RbpA_{Mtb}^{72-111}$ **(Chapter 2 Figure 4e)**. One possible explanation is that the truncation of RbpA NTT and CD antagonizes full-length transcription. To begin dissecting whether RbpA NTT promotes full-length transcription, we compared *M. tuberculosis* RNAP-σ^A activity on the *M. tuberculosis* *rrn*AP3 promoter in the presence of $RbpA_{Mtb}^{WT}$ to that in the presence of $RbpA_{Mtb}^{26-111}$. Similar to previous reports, we found that $RbpA_{Mtb}^{WT}$ increases full-length transcription from *rrn*AP3 **(Chapter 3 Figure 5)**[108]. In contrast, full-length transcription in the presence of $RbpA_{Mtb}^{26-111}$ was attenuated compared to in the presence of $RbpA_{Mtb}^{WT}$, showing that RbpA NTT facilitates full-length transcription.

During logarithmic growth, σ^B regulates approximately 10 percent of *M. smegmatis* genes in a mostly RbpA independent manner

To better understand the role of σ^B in *M. smegmatis*, we deleted *sigB* and compared the transcriptomes of *M. smegmatis* Δ*sigB* and *M. smegmatis* with *sigB* intact during mid-logarithmic growth with RNA-seq. A total of 585 genes were significantly upregulated or downregulated at least 2-fold in *M. smegmatis* Δ*sigB* **(Chapter 4 Figure 1)**. Among these 585 genes, 420 were downregulated at least 2-fold indicating that σ^B contributes to *M. smegmatis* transcription during logarithmic growth mostly through activating transcription. RbpA interacts with both σ^A and σ^B but it is unclear how RbpA impacts the function of σ^A and σ^B *in vivo* [107,108]. To better understand the extent to which σ^B regulated transcription requires interaction with RbpA, we compared the transcriptomes of *M. smegmatis ΔsigB* with *M. smegmatis* $RbpA_{Mtb}^{R88A}$.

The R88A substitution in RbpA SID weakens RbpA's interaction with both σ^A and σ^B and therefore this comparison between *M. smegmatis ΔsigB* and *M. smegmatis* $RbpA_{Mtb}^{R88A}$ allows us to identify which genes are regulated by σ^B in a RbpA independent manner and which genes are regulated in a RbpA dependent manner **(Chapter 2 Figure 2)**[125]. Specifically, genes that are differentially expressed only in *M. smegmatis ΔsigB* are interpreted as genes that are σ^B regulated in a RbpA independent manner and genes that differentially expressed in both strains are considered genes that are σ^B regulated in a RbpA dependent manner. Of the 585 differentially expressed genes in *M. smegmatis ΔsigB* only 23 are also differentially expressed in *M. smegmatis* $RbpA_{Mtb}^{R88A}$ in the same direction indicating that transcription of the σ^B regulon occurs in a mostly RbpA independent fashion **(Chapter 4 Figure 1)**.

20 of the 30 most downregulated genes in *M. smegmatis* encode small gene products

Most of the differentially expressed genes in *M. smegmatis ΔsigB* are downregulated and among the most downregulated genes is a group of small genes **(Chapter 4 Figure 2)**. Little is known about the function of the small genes including whether or not they encode small non-translated RNAs or small proteins. If these σ^B regulated genes are translated, 16 of the 20 would be highly acidic or basic with an isoelectric point less than 4 or greater than 8. Most of the genes do not have known orthologues in other bacterial species suggesting that the proteins are *M. smegmatis* specific or comprise one or more novel classes of proteins.

***M. smegmatis sigB* is made synthetically essential by point mutations in RbpA BL and SID**

Experiments that were designed to differentiate the functions of RbpA bound σ^A and RbpA bound σ^B required engineering *M. smegmatis ΔsigB* strains with point mutations in RbpA BL and SID. So far, we have used three different approaches using a variety of genetic tools to try and obtain either of these two *M. smegmatis* strains and have been unsuccessful **(Chapter 4**

Figure 3). The inability to engineer these strains through various approaches supports the conclusion that *sigB* is synthetically essential when RbpA BL and SID activities are limited.

Open questions

How do other bacterial RNAPs carry out whatever essential function(s) RbpA performs in mycobacteria?

RbpA is only conserved within the *Actinobacteria* phylum, which raises the question of how RNAPs from other bacterial phyla accomplish the tasks performed by RbpA. Some insight into this question has been gained through the comparison of the *M. tuberculosis* RNAP-σ^A holoenzyme and the *E. coli* RNAP-σ^{70} holoenzyme *in vitro* RP_o stability [64,65]. This comparison showed that RbpA increases *M. tuberculosis* RNAP-σ^A holoenzyme RP_o stability and that RbpA is required for the *M. tuberculosis* RNAP-σ^A holoenzyme to reach the same level of RP_o stabilization achieved by the *E. coli* RNAP-σ^{70} [65]. This comparison also indicates that *E. coli*'s housekeeping holoenzyme has one or more RP_o stabilizing features that the *M. tuberculosis* RNAP-σ^A holoenzyme does not have. Characterization of the *E. coli* RNAP-σ^{70} holoenzyme has identified structural regions involved in RP_o stabilization and one of these RP_o stabilizing structural regions is a lineage specific insertion unique to γ-Proteobacteria called the β' sequence insertion 3 (β' SI3)[5]. The similar effect that RbpA and β' SI3 have on increasing RP_o stabilization lends itself to the possibility that RbpA is essential because the *M. tuberculosis* RNAP-σ^A holoenzyme does not have the β' SI3 and *Actinobacteria* have evolved an alternative RbpA dependent solution for stabilizing RP_o. However, even if this turns out to be true the exclusivity of RbpA among Actinobacteria and β' SI3 among γ-proteobacteria indicates that there are still additional RP_o stabilizing mechanisms among the bacterial phyla that lack both RbpA and β' SI3.

RbpA might also be essential for a different reason. One strategy that could potentially shed light on RbpA's essential function is to further explore the functions of bacterial RNAP lineage specific insertions such as the γ-Proteobacteria β' SI3 that connect conserved regions of the RNAP [6,7]. There are a limited number of lineage specific insertions, and one could engineer a cohort of *M. tuberculosis* strains where each strain has a *M. tuberculosis* RNAP-σ^A holoenzyme containing one lineage specific insertion. One could then attempt to delete RbpA from these strains (if these *M. tuberculosis* RNAP-σ^A holoenzyme containing lineage specific insertions are viable) to identify which lineage specific insertions are functionally linked to RbpA.

RbpA could also be essential due to the presence rather than absence of a *M. tuberculosis* RNAP-σ^A holoenzyme feature. Some of these features are discussed in the chapter 1 introduction, including the *Actinobacteria* lineage specific RNAP β' insertion, the N-terminal extension attached to σ^A region 1.1 and the other essential RNAP binding protein CarD. To my knowledge it is unknown whether the *Actinobacteria* lineage specific RNAP β' insertion or σ^A region 1.1 tail are essential in mycobacteria. If these features are not essential one could engineer *M. tuberculosis* strains lacking these structures and then try deleting *rbpA* to test whether RbpA is required because of these structures. CarD is essential and therefore more advanced genetic approaches would be required to test the hypothesis that RbpA is essential because of the existence of CarD. However, a potential starting point could be attempting to delete *rbpA* in *M. tuberculosis* strains encoding CarD point mutations, such the R25E and K125A that limit different CarD functions [94].

Comparisons of *M. tuberculosis* and *E. coli* promoters show that *M. tuberculosis* promoter architecture is less stringently defined compared to *E. coli* [10,90]. The prevalence of multiple promoter sequence elements such as the -35 element and UP element is lower in *M.*

tuberculosis promoters and there is more flexibility in nucleotide identity within *M. tuberculosis* promoter sequence elements such as the -10 element where the consensus sequence is defined as 5' – TANNNT – 3' compared to 5' – TATAAT – 3' in *E. coli* [10,90]. These comparisons raise the question of what mechanisms might compensate for the lack of promoter architecture in *M. tuberculosis.* One hypothesis is that *M. tuberculosis* requires less promoter sequence stringency because RNAP binding factors like RbpA perform the functions carried out by a particular sequence element in *E. coli*. This hypothesis is supported by data showing that the activity of the *M. tuberculosis* RNAP-σ^A holoenzyme at a promoter can be increased by either RbpA or the addition of an extended -10 element to the promoter, suggesting there is functional overlap between the extended -10 element and RbpA [148,152].

Studies defining the consensus promoter sequence elements in *M. tuberculosis* have for the most part focused on the promoter sequence elements that are important in *E. coli*. Though informative, these analyses potentially miss promoter sequence elements found in *M. tuberculosis* that are absent in *E. coli*. Furthermore, RbpA could be utilizing an undiscovered promoter sequence element. Future chromatin immunoprecipitation sequencing (ChIP-seq) experiments that determine which promoters RbpA localizes to in conjunction with our RNA-sequencing experiments will more clearly define the RbpA regulon and allow us to evaluate whether there is a RbpA specific promoter architecture and if so, what sequence elements constitute this architecture.

Why does RbpA bind both σ^A and σ^B?

Structural characterization of RbpA binding to the *M. tuberculosis* RNAP-σ^A holoenzyme has revealed which RbpA residues interact with which σ^A residues [22,67,106]. A comparison of σ^A and σ^B amino acid sequences shows that the σ^A residues involved in the interaction with RbpA

are mostly conserved in σ^B but not conserved in the other *M. tuberculosis* σ factors [22]. Despite knowing how RbpA interacts with σ^A and σ^B, there are still many open questions centered around when and why RbpA interacts with each σ factor. For instance, we do not know how much RbpA bound or unbound σ^A or σ^B exists in the bacterium and we do not know the amount of functional overlap, if any, between σ^A, σ^B, RbpA bound σ^A and RbpA bound σ^B.

We have made some progress in understanding the roles of RbpA bound or unbound σ^A and σ^B. Our co-IP experiments **(Chapter 2 Figure 2)** show that RbpA is bound to both σ^A and σ^B during *M. smegmatis* logarithmic growth and that RbpA's interaction with at least one of the two σ factors is in the context of the RNAP due to RbpA's interaction with σ being required for RbpA co-IP of core RNAP [125]. However, these experiments did not determine how much of the total cellular σ^A or σ^B is bound to RbpA and whether RbpA interacts with both σ^A and σ^B in the context of the RNAP. One way to address this question in future work would be to compare the levels of RbpA and core RNAP co-IPed by σ^A and σ^B.

Another experiment that has shed some light on RbpA bound versus unbound σ^A and σ^B functions is our RNA-seq analysis of the *M. smegmatis* σ^B regulon during logarithmic growth. The results suggest that σ^B contributes to transcription during logarithmic growth along with RbpA bound or unbound σ^A. A comparison of the transcriptomes from *M. smegmatis ΔsigB* and *M. smegmatis* $RbpA_{Mtb}^{R88A}$ shows that σ^B regulation during logarithmic growth occurs in a RbpA independent manner based on the lack of overlap in genes that are differentially expressed when σ^B is lost and when σ^B interaction with RbpA is weakened **(Chapter 4 Figure 1)**. To better define RbpA dependent and independent σ^A and σ^B functions in the future we have begun the process of engineering two *M. smegmatis* strains that have a point mutation in either σ^A or σ^B at a residue that is predicted to form a polar interaction with RbpA R88. If the σ^A and σ^B point

mutants are viable and weaken σ^A or σ^B interaction with RbpA without impacting any other σ^A or σ^B activities, these strains could prove useful in separating out RbpA bound and unbound σ^A and σ^B functions.

Why is *sigA* essential and *sigB* non-essential in mycobacteria?

SigA encodes the essential principal housekeeping σ factor while *sigB* encodes the non-essential principal-like σ factor. *In vitro* analysis of *M. tuberculosis* RNAP-σ^A holoenzyme and *M. tuberculosis* RNAP-σ^B holoenzyme has shown that at the *M. tuberculosis* housekeeping *rrn*AP3 promoter, the activity of RNAP-σ^A holoenzyme is greater than that of RNAP-σ^B holoenzyme providing an explanation as to why *sigA* is essential and *sigB* is not (Drake Jensen and Eric Galburt, unpublished data). However, to my knowledge it is not clear why σ^A promotes higher *M. tuberculosis* holoenzyme activity compared to σ^B. Overall the structure of σ^A and σ^B are similar with the exception of the long σ^A region 1.1 N-terminal tail extension. The enigmatic structure and function of the σ^A region 1.1 N-terminal extension have eluded characterization but is likely to at least in part explain why *sigA* is essential. We have begun the process of engineering a *M. tuberculosis sigA* construct lacking the region 1.1 N-terminal extension and plan to also engineer a *M. tuberculosis sigB* construct with the σ^A region 1.1 N-terminal extension to address whether this structural region underpins *sigA*'s status as an essential σ factor.

What are the determinants of RbpA transcriptional activation versus repression?

RNA-seq analysis of *M. smegmatis* RbpA mutants suggests that RbpA activates transcription at some promoters and represses transcription at other promoters (**Chapter 2 Figure 4**). A thermodynamics based model for how a factor can both activate and repress transcription has been developed by Eric Galburt's lab, which hypothesizes that transcriptional

flux is a function of basal rate constants during transcription initiation that vary among promoters [135]. Moreover, depending on the basal rate constants of a given promoter, which rate constants a factor such as RbpA affects and how much the factor changes the basal rate constants, the transcriptional flux can be either increased or decreased by a factor which results in either activation or repression of transcription. The Galburt lab and our lab have together determined that RbpA affects rate constants of promoter melting and promoter escape during transcription initiation on the *M. tuberculosis rrn*AP3 promoter **(Chapter 2, Figure 3)**[65,69,125]. Based on these findings, we know that the basal rate constants of promoter melting and promoter escape are determinants of whether RbpA activates or represses transcription. However, we have much to learn about what determines the basal rate constants of promoter melting and promoter escape for *M. tuberculosis* promoters with many factors to consider such as promoter sequence elements, the presence or absence of other factors such as CarD and chromosomal architecture. This question is further complicated by RbpA's four structural domains each having its own effect on promoter melting and promoter escape as indicated by the finding that RbpA NTT and CD antagonize RP_o stability, while the BL and SID are required for RbpA's RP_o stabilizing activity **(Chapter 2 Figure 3)**[125]. Lastly, we have not yet fully explored whether RbpA has additional functions that could impact transcriptional flux via transcription initiation or one of the other steps required for full-length transcription.

References

1. Steitz TA. A mechanism for all polymerases. *Nat News Rev*. 1998;391. doi:10.1080/08940889508602814
2. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of thermus aquaticus core RNA polymerase at 3.3 å resolution. *Cell*. 1999;98(6):811-824.

doi:10.1016/S0092-8674(00)81515-9

3. Cramer P, Bushnell DA, Fu J, et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science (80-).* 2000;288(5466):640-649. doi:10.1126/science.288.5466.640

4. Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*. 2008;451(7180):851-854. doi:10.1038/nature06530

5. Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record MT. E. coli RNA Polymerase Determinants of Open Complex Lifetime and Structure. *J Mol Biol*. 2015;427(15):2435-2450. doi:10.1016/j.jmb.2015.05.024

6. Lane WJ, Darst SA. Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. *J Mol Biol*. 2010;395(4):686-704. doi:10.1016/j.jmb.2009.10.063

7. Lane WJ, Darst SA. Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis. *J Mol Biol*. 2010;395(4):671-685. doi:10.1016/j.jmb.2009.10.062

8. Yu Z, Feng Y, Chatterjee S, et al. Structural Basis of Transcription Initiation. *Science (80-).* 2008;338(November):1076-1081. doi:10.1002/9780470048672.wecb606

9. Murakami KS. X-ray crystal structure of escherichia coli RNA polymerase σ70 holoenzyme. *J Biol Chem*. 2013;288(13):9126-9134. doi:10.1074/jbc.M112.430900

10. Hubin EA, Lilic M, Darst SA, Campbell EA. Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nat Commun*. 2017;8(May):1-12. doi:10.1038/ncomms16072

11. Tan L, Wiesler S, Trzaska D, Carney HC, Weinzierl ROJ. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol*. 2008;7(10). doi:10.1186/jbiol98

12. Mazumder A, Lin M, Kapanidis AN, Ebright RH. Closing and opening of the RNA polymerase trigger loop. *Proc Natl Acad Sci U S A*. 2020;117(27):15642-15649. doi:10.1073/pnas.1920427117

13. Sosunov V, Zorov S, Sosunova E, et al. The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res*. 2005;33(13):4202-4211. doi:10.1093/nar/gki688

14. Zaychikov E, Martin E, Denissova L, et al. Mapping of Catalytic Residues in the RNA Polymerase Active Center. *Science (80-)*. 1996;273(5271):107-109.

15. Burgess RR, Travers AA, Dunn JJ, Bautz EKF. Factor stimulating transcription by RNA polymerase. *Nature*. 1969;221(5175):43-46. doi:10.1038/221043a0

16. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. MiST 3.0: An updated microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic Acids Res*. 2020;48(1):459-464. doi:10.1093/nar/gkz988

17. Danson AE, Jovanovic M, Buck M, Zhang X. Mechanisms of σ54-Dependent Transcription Initiation and Regulation. *J Mol Biol*. 2019;431(20):3960-3974. doi:10.1016/j.jmb.2019.04.022

18. Flentie K, Garner AL, Stallings CL. Mycobacterium tuberculosis transcription machinery: Ready to respond to host attacks. *J Bacteriol*. 2016;198(9):1360-1373. doi:10.1128/JB.00935-15

19. Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: A historical, structural, and genomic perspective. *Annu Rev Microbiol*. 2014;68:357-376. doi:10.1146/annurev-micro-092412-155737

20. Lonetto MA, Gribskov M, Gross CA. The σ70 Family : Sequence Conservation and

Evolutionary Relationships. *J Bacteriol*. 1992;174(12):3843-3849.

21. Lonetto MA, Brown KL, Rudd KE, Buttner MJ. Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A*. 1994;91(16):7573-7577. doi:10.1073/pnas.91.16.7573

22. Hubin EA, Tabib-Salazar A, Humphrey LJ, et al. Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc Natl Acad Sci U S A*. 2015;112(23):7171-7176. doi:10.1073/pnas.1504942112

23. Sorenson MK, Ray SS, Darst SA. Crystal structure of the flagellar σ/anti-σ complex σ28/FlgM reveals an intact σ factor in an inactive conformation. *Mol Cell*. 2004;14(1):127-138. doi:10.1016/S1097-2765(04)00150-9

24. Sorenson MK, Darst SA. Disulfide cross-linking indicates that FIgM-bound and free σ28 adopt similar conformations. *Proc Natl Acad Sci U S A*. 2006;103(45):16722-16727. doi:10.1073/pnas.0606482103

25. Mekler V, Kortkhonjia E, Mukhopadhyay J, et al. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell*. 2002;108(5):599-614. doi:10.1016/S0092-8674(02)00667-0

26. Bae B, Davis E, Brown D, Campbell EA, Wigneshweraraj S, Darst SA. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. *Proc Natl Acad Sci U S A*. 2013;110(49):19772-19777. doi:10.1073/pnas.1314576110

27. Pletnev P, Pupov D, Pshanichnaya L, et al. Rewiring of growth-dependent transcription regulation by a point mutation in region 1.1 of the housekeeping σ factor. *Nucleic Acids Res*. 2020;48(19):10802-10819. doi:10.1093/nar/gkaa798

28. Kulbachinskiy A, Mustaev A. Region 3.2 of the σ subunit contributes to the binding of the 3′-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J Biol Chem*. 2006;281(27):18273-18276. doi:10.1074/jbc.C600060200

29. Pupov D, Kuzin I, Bass I, Kulbachinskiy A. Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res*. 2014;42(7):4494-4504. doi:10.1093/nar/gkt1384

30. Cashel M, Hsu LM, Hernandez VJ. Changes in conserved region 3 of Escherichia coli σ70 reduce abortive transcription and enhance promoter escape. *J Biol Chem*. 2003;278(8):5539-5547. doi:10.1074/jbc.M211430200

31. Pupov D, Petushkov I, Esyunina D, Murakami KS, Kulbachinskiy A. Region 3.2 of the factor controls the stability of rRNA promoter complexes and potentiates their repression by DksA. *Nucleic Acids Res*. 2018;46(21):11477-11487. doi:10.1093/nar/gky919

32. Callaci S, Heyduk E, Heyduk T. Conformational changes of Escherichia coli RNA polymerase σ70 factor induced by binding to the core enzyme. *J Biol Chem*. 1998;273(49):32995-33001. doi:10.1074/jbc.273.49.32995

33. Callaci S, Heyduk E, Heyduk T, Doisy EA. Core RNA polymerase from E. coli induces a major change in the domain arrangement of the σ70 subunit. *Mol Cell*. 1999;3(2):229-238. doi:10.1016/S1097-2765(00)80313-5

34. Murakami KS, Masuda S, Darst SA. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science (80-)*. 2002;296(5571):1280-1284. doi:10.1126/science.1069594

35. Maeda H, Fujita N, Ishihama A. Competition among seven Escherichia coli σ subunits:

Relative binding affinities to the core RNA polymerase. *Nucleic Acids Res*. 2000;28(18):3497-3503. doi:10.1093/nar/28.18.3497

36. Arthur TM, Anthony LC, Burgess RR. Mutational analysis of β260-309, a σ70 binding site located on Escherichia coli core RNA polymerase. *J Biol Chem*. 2000;275(30):23113-23119. doi:10.1074/jbc.M002040200

37. Young BA, Anthony LC, Gruber TM, et al. A coiled-coil from the RNA polymerase β′ subunit allosterically induces selective nontemplate strand binding by σ70. *Cell*. 2001;105(7):935-944. doi:10.1016/S0092-8674(01)00398-1

38. Burgess RR, Anthony L. How sigma docks to RNA polymerase and what sigma does. *Curr Opin Microbiol*. 2001;4(2):126-131. doi:10.1016/S1369-5274(00)00177-6

39. Li L, Fang C, Zhuang N, Wang T, Zhang Y. Structural basis for transcription initiation by bacterial ECF σ factors. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-09096-y

40. Lin W, Mandal S, Degen D, et al. Structural basis of ECF-σ-factor-dependent transcription initiation. *Nat Commun*. 2019;10(1):1-14. doi:10.1038/s41467-019-08443-3

41. Pribnow D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A*. 1975;72(3):784-788. doi:10.1073/pnas.72.3.784

42. Schaller H, Gray C, Herrmann K. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc Natl Acad Sci U S A*. 1975;72(2):737-741. doi:10.1073/pnas.72.2.737

43. Maniatis T, Ptashne M, Backman K, et al. Recognition sequences of repressor and polymerase in the operators of bacteriophage lambda. *Cell*. 1975;5(2):109-113. doi:10.1016/0092-8674(75)90018-5

44. Hawley DK, Mcclure WR, Limited IRLP. Compilation and analysis of Escherichia coli

promoter DNA sequences. *Nucleic Acids Res*. 1983;11(8).

45. Sclavi B, Zaychikov E, Rogozina A, Walther F, Buckle M, Heumann H. 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*. 2005;102(13):4706-4711. doi:10.1073/pnas.0408218102

46. Campbell EA, Muzzin O, Chlenov M, et al. Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol Cell*. 2002;9(3):527-539. doi:10.1016/S1097-2765(02)00470-7

47. Travers AA. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J Bacteriol*. 1980;141(2):973-976. doi:10.1128/jb.141.2.973-976.1980

48. Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA Promoter Regulation by Nonoptimal Binding of σ Region 1.2: An Additional Recognition Element for RNA Polymerase. *Cell*. 2006;125(6):1069-1082. doi:10.1016/j.cell.2006.04.034

49. Feklistov A, Barinova N, Sevostyanova A, et al. A Basal Promoter Element Recognized by Free RNA Polymerase σ Subunit Determines Promoter Recognition by RNA Polymerase Holoenzyme. *Mol Cell*. 2006;23(1):97-107. doi:10.1016/j.molcel.2006.06.010

50. Ponnambalam S, Webster C, Bingham A, Busby S. Transcription initiation at the Escherichia coli galactose operon promoters in the absence of the normal -35 region sequences. *J Biol Chem*. 1986;261(34):16043-16048. doi:10.1016/s0021-9258(18)66673-6

51. Keilty S, Rosenberg M. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J Biol Chem*. 1987;262(13):6389-6395. doi:10.1016/s0021-9258(18)45582-2

52. Barne KA, Bown JA, Busby SJW, Minchin SD. Region 2.5 of the Escherichia coli RNA polymerase σ70 subunit is responsible for the recognition of the "extended -10" motif at promoters. *EMBO J*. 1997;16(13):4034-4040. doi:10.1093/emboj/16.13.4034

53. Sanderson A, Mitchell JE, Minchin SD, Busby SJW. Substitutions in the Escherichia coli RNA polymerase σ70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett*. 2003;544(1-3):199-205. doi:10.1016/S0014-5793(03)00500-3

54. Ross W, Gosink KK, Salomon J, et al. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science (80-)*. 1993;262(5138):1407-1413. doi:10.1126/science.8248780

55. Davis CA, Bingman CA, Landick R, Record MT, Saecker RM. 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*. 2007;104(19):7833-7838. doi:10.1073/pnas.0609888104

56. Naryshkin N, Revyakin A, Kim Y, Mekler V, Ebright RH. Structural organization of the RNA polymerase-promoter open complex. *Cell*. 2000;101(6):601-611. doi:10.1016/S0092-8674(00)80872-7

57. Spassky A, Kirkegaard K, Buc H. Changes in the DNA Structure of the lac UV5 Promoter during Formation of an Open Complex with Escherichia coli RNA Polymerase. *Biochemistry*. 1985;24(11):2723-2731. doi:10.1021/bi00332a019

58. Kovacic RT. The 0 degree C closed complexes between Escherichia coli RNA polymerase and two promoters, T7-A3 and lacUV5. *J Biol Chem*. 1987;262(28):13654-13661. doi:10.1016/s0021-9258(19)76477-1

59. Craig ML, Suh WC, Thomas Record M. HO. and DNase I Probing of Eσ70 RNA

Polymerase-λPR Promoter Open Complexes: Mg2+ Binding and Its Structural Consequences at the Transcription Start Site. *Biochemistry*. 1995;34(48):15624-15632. doi:10.1021/bi00048a004

60. Saecker RM, Tsodikov O V., McQuade KL, Schlax PE, Capp MW, Thomas Record M. Kinetic studies and structural models of the association of E. coli σ70 RNA polymerase with the λPR promoter: Large scale conformational changes in forming the kinetically significant intermediates. *J Mol Biol*. 2002;319(3):649-671. doi:10.1016/S0022-2836(02)00293-0

61. Saecker RM, Record MT, Dehaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol*. 2011;412(5):754-771. doi:10.1016/j.jmb.2011.01.018

62. Schickor P, Metzger W, Werel W, Lederer H, Heumann H. Topography of intermediates in transcription initiation of E. coli. *EMBO J*. 1990;9(7):2215-2220. doi:10.1002/j.1460-2075.1990.tb07391.x

63. Mekler V, Minakhin L, Kuznedelov K, Mukhamedyarov D, Severinov K. RNA polymerase-promoter interactions determining different stability of the Escherichia coli and Thermus aquaticus transcription initiation complexes. *Nucleic Acids Res*. 2012;40(22):11352-11362. doi:10.1093/nar/gks973

64. Rammohan J, Ruiz Manzano a., Garner a. L, Stallings CL, Galburt E a. CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Res*. Published online 2015:1-14. doi:10.1093/nar/gkv078

65. Rammohan J, Ruiz Manzano A, Garner AL, Prusa J, Stallings CL, Galburt EA.

Cooperative stabilization of Mycobacterium tuberculosis rrnAP3 promoter open complexes by RbpA and CarD. *Nucleic Acids Res*. 2016;44(15):7304-7313. doi:10.1093/nar/gkw577

66. Davis E, Chen J, Leon K, Darst SA, Campbell EA. Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res*. 2015;43(1):433-445. doi:10.1093/nar/gku1231

67. Boyaci H, Chen J, Lilic M, et al. Fidaxomicin jams mycobacterium tuberculosis RNA polymerase motions needed for initiation via RBPA contacts. *Elife*. 2018;7:1-19. doi:10.7554/eLife.34823

68. Boyaci H, Chen J, Jansen R, Darst SA, Campbell EA. Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature*. 2019;565(7739):382-385. doi:10.1038/s41586-018-0840-5

69. Jensen D, Manzano AR, Rammohan J, Stallings CL, Galburt EA. CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. *Nucleic Acids Res*. 2019;47(13):6685-6698. doi:10.1093/nar/gkz449

70. Chen J, Chiu C, Gopalkrishnan S, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol Cell*. 2020;78(2):275-288.e6. doi:10.1016/j.molcel.2020.02.017

71. Schroeder LA, Gries TJ, Saecker RM, Record MT, Harris ME, deHaseth PL. Evidence for a Tyrosine-Adenine Stacking Interaction and for a Short-lived Open Intermediate Subsequent to Initial Binding of Escherichia coli RNA Polymerase to Promoter DNA. *J Mol Biol*. 2009;385(2):339-349. doi:10.1016/j.jmb.2008.10.023

72. Karpen ME, deHaseth PL. Base flipping in open complex formation at bacterial

promoters. *Biomolecules*. 2015;5(2):668-678. doi:10.3390/biom5020668

73. Feklistov A, Darst SA. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell*. 2011;147(6):1257-1269. doi:10.1016/j.cell.2011.10.041

74. Feklistov A, Bae B, Hauver J, et al. RNA polymerase motions during promoter melting. *Science (80-)*. 2017;866(May):863-866.

75. Drennan A, Kraemer M, Capp M, et al. Polymerase in Transcription Initiation. Published online 2012.

76. Mekler V, Minakhin L, Borukhov S, Mustaev A, Severinov K. Coupling of downstream RNA polymerase-promoter interactions with formation of catalytically competent transcription initiation complex. *J Mol Biol*. 2014;426(24):3973-3984. doi:10.1016/j.jmb.2014.10.005

77. Achillefs C, Kapanidis N, Margeat E, et al. Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism: Single-molecule fluorescence-resonance-energy-transfer experiments establish that initial transcription proceeds through a " scrunching " mechanism, in which RN. *Science (80-)*. 2006;314(5802):1144-1147. doi:10.1126/science.1131399.INITIAL

78. Revyakin A, Liu C, Ebright RH, Strick TR. Abortive Initiation and Productive Initiation By Rna. *Science (80-)*. 2006;314(5802):1139-1143. doi:10.1126/science.1131398.ABORTIVE

79. Winkelman JT, Winkelman BT, Boyce J, et al. Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol Cell*. 2015;59(5):768-780. doi:10.1016/j.molcel.2015.06.037

80. Carpousis AJ, Gralla JD. Cycling of Ribonucleic Acid Polymerase to Produce Oligonucleotides During Initiation in Vitro at the Lac UV5 Promoter. *Biochemistry*. 1980;19(14):3245-3253. doi:10.1021/bi00555a023

81. Ko J, Heyduk T. Kinetics of promoter escape by bacterial RNA polymerase: Effects of promoter contacts and transcription bubble collapse. *Biochem J*. 2014;463(1):135-144. doi:10.1042/BJ20140179

82. Vo N V., Hsu LM, Kane CM, Chamberlin MJ. In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 2. Formation and characterization of two distinct classes of initial transcribing complexes. *Biochemistry*. 2003;42(13):3787-3797. doi:10.1021/bi0269613

83. Zuo Y, Steitz TA. Crystal structures of the e.coli transcription initiation complexes with a complete bubble. *Mol Cell*. 2015;58(3):534-540. doi:10.1016/j.molcel.2015.03.010

84. Belogurov GA, Artsimovitch I. The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J Mol Biol*. 2019;431(20):3975-4006. doi:10.1016/j.jmb.2019.05.042

85. Mustaev A, Roberts J, Gottesman M. Transcription elongation. *Transcription*. 2017;8(3):150-161. doi:10.1080/21541264.2017.1289294

86. Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem*. 2016;85:319-347. doi:10.1146/annurev-biochem-060815-014844

87. Roberts JW. Mechanisms of Bacterial Transcription Termination. *J Mol Biol*. 2019;431(20):4030-4039. doi:10.1016/j.jmb.2019.04.003

88. Lin W, Das K, Degen D, et al. Structural Basis of Transcription Inhibition by Fidaxomicin

(Lipiarmycin A3). *Mol Cell*. 2018;70(1):60-71.e15. doi:10.1016/j.molcel.2018.02.026

89. Manganelli R. Sigma factors: Key molecules in Mycobacterium tuberculosis physiology and virulence. *Mol Genet Mycobact*. Published online 2015:135-160. doi:10.1128/9781555818845.ch7

90. Cortes T, Schubert OT, Rose G, et al. Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in Mycobacterium tuberculosis. *Cell Rep*. 2013;5(4):1121-1131. doi:10.1016/j.celrep.2013.10.031

91. Shell SS, Wang J, Lapierre P, et al. Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genet*. 2015;11(11):1-31. doi:10.1371/journal.pgen.1005641

92. Newton-Foot M, Gey Van Pittius NC. The complex architecture of mycobacterial promoters. *Tuberculosis*. 2013;93(1):60-74. doi:10.1016/j.tube.2012.08.003

93. Stallings CL, Stephanou NC, Chu L, Hochschild A, Nickels BE, Glickman MS. CarD Is an Essential Regulator of rRNA Transcription Required for Mycobacterium tuberculosis Persistence. *Cell*. 2009;138(1):146-159. doi:10.1016/j.cell.2009.04.041

94. Garner AL, Weiss LA, Manzano AR, Galburt EA, Stallings CL. CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. *Mol Microbiol*. 2014;93(4):682-697. doi:10.1111/mmi.12681

95. Weiss LA, Harrison PG, Nickels BE, et al. Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. *J Bacteriol*. 2012;194(20):5621-5631. doi:10.1128/JB.00879-12

96. Bae B, Chen J, Davis E, Leon K, Darst SA, Campbell EA. CarD uses a minor groove

wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife*. 2015;4(September 2015):1-19. doi:10.7554/eLife.08505

97. Srivastava DB, Leon K, Osmundson J, et al. Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A*. 2013;110(31):12619-12624. doi:10.1073/pnas.1308270110

98. Garner AL, Rammohan J, Huynh JP, et al. Effects of increasing the affinity of CarD for RNA polymerase on Mycobacterium tuberculosis growth, rRNA transcription, and virulence. *J Bacteriol*. 2017;199(4). doi:10.1128/JB.00698-16

99. Zhu DX, Garner AL, Galburt EA, Stallings CL. CarD contributes to diverse gene expression outcomes throughout the genome of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*. 2019;116(27):13573-13581. doi:10.1073/pnas.1900176116

100. Landick R, Krek A, Glickman MS, Socci ND, Stallings CL. Genome-wide mapping of the distribution of CarD, RNAP σA, and RNAP β on the Mycobacterium smegmatis chromosome using chromatin immunoprecipitation sequencing. *Genomics Data*. 2014;2:110-113. doi:10.1016/j.gdata.2014.05.012

101. Paget MSB, Molle V, Cohen G, Aharonowitz Y, Buttner MJ. Defining the disulphide stress response in Streptomyces coelicolor A3(2): Identification of the σR regulon. *Mol Microbiol*. 2001;42(4):1007-1020. doi:10.1046/j.1365-2958.2001.02675.x

102. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol Microbiol*. 2002;43(3):717-731. doi:10.1046/j.1365-2958.2002.02779.x

103. Stewart GR, Wernisch L, Stabler R, et al. Dissection of the heat-shock response in

Mycobacterium tuberculosis using mutants and microarrays. *Microbiology*. 2002;148(10):3129-3138. doi:10.1099/00221287-148-10-3129

104. Murphy DJ, Brown JR. Identification of gene targets against dormant phase Mycobacterium tuberculosis infections. *BMC Infect Dis*. 2007;7:1-16. doi:10.1186/1471-2334-7-84

105. Provvedi R, Boldrin F, Falciani F, Palù G, Manganelli R. Global transcriptional response to vancomycin in Mycobacterium tuberculosis. *Microbiology*. 2009;155(4):1093-1102. doi:10.1099/mic.0.024802-0

106. Hubin EA, Fay A, Xu C, et al. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife*. 2017;6:1-40. doi:10.7554/eLife.22520

107. Hu Y, Morichaud Z, Chen S, Leonetti JP, Brodolin K. Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the σ A-containing RNA polymerase holoenzyme. *Nucleic Acids Res*. 2012;40(14):6547-6557. doi:10.1093/nar/gks346

108. Hu Y, Morichaud Z, Perumal AS, Roquet-Baneres F, Brodolin K. Mycobacterium RbpA cooperates with the stress-response σB subunit of RNA polymerase in promoter DNA unwinding. *Nucleic Acids Res*. 2014;42(16):10399-10408. doi:10.1093/nar/gku742

109. Bortoluzzi A, Muskett FW, Waters LC, et al. Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. *J Biol Chem*. 2013;288(20):14438-14450. doi:10.1074/jbc.M113.459883

110. Tabib-Salazar A, Liu B, Doughty P, et al. The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucleic Acids Res*.

2013;41(11):5679-5691. doi:10.1093/nar/gkt277

111. Pashley CA, Parish T. Efficient switching of mycobacteriophage L5-based integrating plasmids in Mycobacterium tuberculosis. *FEMS Microbiol Lett*. 2003;229(2):211-215. doi:10.1016/S0378-1097(03)00823-1

112. Barkan D, Stallings CL, Glickman MS. An improved counterselectable marker system for mycobacterial recombination using galK and 2-deoxy-galactose. *Gene*. 2011;470(1-2):31-36. doi:10.1016/j.gene.2010.09.005

113. Banerjee R, Rudra P, Prajapati RK, Sengupta S, Mukhopadhyay J. Optimization of recombinant Mycobacterium tuberculosis RNA polymerase expression and purification. *Tuberculosis*. 2014;94(4):397-404. doi:10.1016/j.tube.2014.03.008

114. Banerjee R, Rudra P, Saha A, Mukhopadhyay J. Recombinant reporter assay using transcriptional machinery of Mycobacterium tuberculosis. *J Bacteriol*. 2015;197(3):646-653. doi:10.1128/JB.02445-14

115. Rammohan J, Manzano AR, Garner AL, Stallings CL, Galburt EA. CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Res*. 2015;43(6):3272-3285. doi:10.1093/nar/gkv078

116. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

117. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

118. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656

119. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21. doi:10.1186/s13059-014-0550-8

120. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210. doi:10.1093/nar/30.1.207

121. Gonzalez-y-Merchand JA, Colstonl MJ, Cox RA. The rRNA operons of Mycobacterium smegmatis and Mycobacterium tuberculosis: comparison of neighbouring upstream genes. *Microbiology*. 1996;(142):667-674.

122. Albery JW, Knowles JR. Evolution of Enzyme Function and the Development of Catalytic Efficiency. *Biochemistry*. 1976;15(25):5631-5640.

123. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol*. 2016;36(5):662-667. doi:10.1128/mcb.00970-14

124. Raju RM, Unnikrishnan M, Rubin DHF, et al. Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathog*. 2012;8(2). doi:10.1371/journal.ppat.1002511

125. Prusa J, Jensen D, Santiago-Collazo G, et al. Domains within RbpA serve specific functional roles that regulate the expression of distinct mycobacterial gene subsets. *J Bacteriol*. 2018;200(13). doi:10.1128/JB.00690-17

126. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science (80-).* 2001;292(5523):1876-1882. doi:10.1126/science.1059495

127. Kuznedelov K, Minakhin L, Niedziela-Majka A, et al. A role for interaction of the RNA polymerase flap domain with the σ subunit in promoter recognition. *Science (80-).* 2002;295(5556):855-857. doi:10.1126/science.1066303

128. Vishwakarma RK, Brodolin K. The σ Subunit-Remodeling Factors: An Emerging Paradigms of Transcription Regulation. *Front Microbiol*. 2020;11(July):1-9. doi:10.3389/fmicb.2020.01798

129. Dove SL, Darst SA, Hochschild A. Region 4 of σ as a target for transcription regulation. *Mol Microbiol*. 2003;48(4):863-874. doi:10.1046/j.1365-2958.2003.03467.x

130. Toulokhonov I, Landick R. The Role of the Lid Element in Transcription by E. coli RNA Polymerase. *J Mol Biol*. 2006;361(4):644-658. doi:10.1016/j.jmb.2006.06.071

131. Naryshkina T, Kuznedelov K, Severinov K. The Role of the Largest RNA Polymerase Subunit Lid Element in Preventing the Formation of Extended RNA-DNA Hybrid. *J Mol Biol*. 2006;361(4):634-643. doi:10.1016/j.jmb.2006.05.034

132. Shi W, Zhou W, Zhang B, et al. Structural basis of bacterial σ 28 -mediated transcription reveals roles of the RNA polymerase zinc-binding domain . *EMBO J*. 2020;39(14):1-15. doi:10.15252/embj.2020104389

133. Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N. A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J*. 2011;30(18):3766-3775. doi:10.1038/emboj.2011.252

134. Morichaud Z, Chaloin L, Brodolin K. Regions 1.2 and 3.2 of the RNA Polymerase σ Subunit Promote DNA Melting and Attenuate Action of the Antibiotic Lipiarmycin. *J Mol Biol*. 2016;428(2):463-476. doi:10.1016/j.jmb.2015.12.017

135. Galburt EA. The calculation of transcript flux ratios reveals single regulatory mechanisms

capable of activation and repression. *Proc Natl Acad Sci U S A*. 2018;115(50):E11604-E11613. doi:10.1073/pnas.1809454115

136. Gottesman S. Trouble is coming: Signaling pathways that regulate general stress responses in bacteria. *J Biol Chem*. 2019;294(31):11685-11700. doi:10.1074/jbc.REV119.005593

137. Hu Y, Coates ARM. Transcription of two sigma 70 homologue genes, sigA and sigB, in stationary-phase Mycobacterium tuberculosis. *J Bacteriol*. 1999;181(2):469-476. doi:10.1128/jb.181.2.469-476.1999

138. Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I. Differential expression of 10 sigma factor genes in Mycobacterium tuberculosis. *Mol Microbiol*. 1999;31(2):715-724. doi:10.1046/j.1365-2958.1999.01212.x

139. Dutta NK, Mehra S, Kaushal D. A Mycobacterium tuberculosis sigma factor network responds to cell-envelope damage by the promising anti-mycobacterial thioridazine. *PLoS One*. 2010;5(4). doi:10.1371/journal.pone.0010069

140. Fontán PA, Voskuil MI, Gomez M, et al. The Mycobacterium tuberculosis sigma factor σB is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for in vivo growth. *J Bacteriol*. 2009;191(18):5628-5633. doi:10.1128/JB.00510-09

141. Pisu D, Provvedi R, Espinosa DM, et al. The alternative sigma factors SigE and SigB are involved in tolerance and persistence to antitubercular drugs. *Antimicrob Agents Chemother*. 2017;61(11):1-11. doi:10.1128/AAC.01596-17

142. Kurthkoti K, Amin H, Marakalala MJ, et al. The capacity of mycobacterium tuberculosis to survive iron starvation might enable it to persist in iron- deprived microenvironments of

human granulomas. *MBio*. 2017;8(4):1-17. doi:10.1128/mBio.01092-17

143. Hurst-Hess K, Biswas R, Yang Y, Rudra P, Lasek-Nesselquist E, Ghosh P. Mycobacterial SigA and SigB cotranscribe essential housekeeping genes during exponential growth. *MBio*. 2019;10(3):1-17. doi:10.1128/mBio.00273-19

144. Raman S, Song T, Puyang X, Bardarov S, Jacobs J, Husson RN. The alternative sigma factor sigh regulates major components of oxidative and heat stress responses in Mycobacterium tuberculosis. *J Bacteriol*. 2001;183(20):6119-6125. doi:10.1128/JB.183.20.6119-6125.2001

145. He H, Hovey R, Kane J, Singh V, Zahrt TC. Erratum: MprAB is a stress-responsive two-component system that directly regulates expression of sigma factors SigB and SigE in Mycobacterium tuberculosis (Journal of Bacteriology (2006) 188:6 (2138) DOI: 10.1128/JB.188.6.2134-2143.2006). *J Bacteriol*. 2020;202(20):2134-2143. doi:10.1128/JB.00443-20

146. Song T, Song SE, Raman S, Anaya M, Husson RN. Critical role of a single position in the -35 element for promoter recognition by Mycobacterium tuberculosis SigE and SigH. *J Bacteriol*. 2008;190(6):2227-2230. doi:10.1128/JB.01642-07

147. Lange R, Hengge-Aronis R. The cellular concentration of the σ(S) subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes Dev*. 1994;8(13):1600-1612. doi:10.1101/gad.8.13.1600

148. Perumal AS, Vishwakarma RK, Hu Y, Morichaud Z, Brodolin K. RbpA relaxes promoter selectivity of M. tuberculosis RNA polymerase. *Nucleic Acids Res*. 2018;46(19):10106-10118. doi:10.1093/nar/gky714

149. Zhu Y, Mao C, Ge X, Wang Z, Lu P. crossm Characterization of a Minimal Type of

Position in Mycobacteria. 2017;199(21):1-12.

150. Bardarov S, Bardarov S, Pavelka MS, et al. Specialized transduction: An efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. *Microbiology*. 2002;148(10):3007-3017. doi:10.1099/00221287-148-10-3007

151. Duval M, Cossart P. Small bacterial and phagic proteins: an updated view on a rapidly moving field. *Curr Opin Microbiol*. 2017;39:81-88. doi:10.1016/j.mib.2017.09.010

152. Vishwakarma RK, Cao AM, Morichaud Z, Perumal AS, Margeat E, Brodolin K. Single-molecule analysis reveals the mechanism of transcription activation in M. tuberculosis. *Sci Adv*. 2018;4(5):1-9. doi:10.1126/sciadv.aao5498