Spring 5-15-2023

Mechanisms of transcription regulation by CarD in Mycobacterium

Dennis Xiaoqian Zhu

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Plant and Microbial Biosciences

Dissertation Examination Committee:

Christina Stallings, Chair
Petra Levin, Co-Chair
Gautam Dantas
Eric Galburt
Makedonka Mitreva
David Wang

Mechanisms of Transcription Regulation by CarD in *Mycobacterium*

by
Dennis Zhu

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2023

St. Louis, Missouri

Table of Contents

List of Figures

Chapter 2:

Chapter 3:

Chapter 4:

List of Tables

Chapter 1:

Chapter 3:

Chapter 4:

Acknowledgments

First and foremost, I would like to acknowledge my thesis mentor Dr. Christina Stallings. None of the work described in this document would have been possible without her support and guidance both as a scientific mentor and as a role model in life. Christina is an energetic force of nature in all aspects of her life yet seems impossibly grounded and easy to talk to. I am thankful for the laboratory environment that Christina has cultivated, which has felt like an additional family for me. In particular, I would like to acknowledge several people who were present when I first joined the lab – Jerome Prusa, Greg Harrison, and Rachel Kinsella, who taught me everything that I know about microbiology and supported me through my mistakes as a young graduate student. I'd also like to acknowledge the other graduate students in the lab, including Michael Nehls, Skyler Hendrix, Helen Blaine, Erin Wang, and Ananda Rankin, who have provided invigorating scientific conversations and treasured friendships. I'd also like to thank all of the members of my thesis committee, especially my thesis chair Dr. Petra Levin, who is responsible for introducing me to the field of bacteriology and has supported me through chats about science, running, and dogs. I am also grateful for Dr. Eric Galburt, Drake Jensen, and other members of the Galburt lab for their scientific collaboration and helping me struggle through all things biochemical and biophysical. I'd also like to thank my undergraduate research mentor Dr. Paula McSteen, without whom I would not be a scientist today. I am also grateful for the faculty and staff in the Molecular Microbiology Department at Washington University who have created a wonderful and safe learning environment.

I am also grateful for my friends who have made my life vibrant, especially my friends from the Mizzou Running Club and the Furious George volleyball team. Finally, I would like to thank my family, especially my parents Yimin and Ce, my sister Meimei, my brother-in-law Sam.

Graduate school often feels incredibly de-stabilizing, but your love and support has always felt like a comforting shelter from the chaos of life.

Dennis Zhu

Washington University in St. Louis

May 2023

Dedicated to my parents Yimin and Ce, my sister Meimei, and Rosa.

ABSTRACT OF THE DISSERTATION

Mechanisms of transcription regulation by CarD in *Mycobacterium*

by

Dennis Xiaoqian Zhu

Doctor of Philosophy in Biology and Biomedical Sciences

Program in Plant and Microbial Biosciences

Washington University in St. Louis, 2023

Professor Christina L. Stallings, Chair

Professor Petra A. Levin, Co-chair

Mycobacterium tuberculosis (*Mtb*) is an important human pathogen that is responsible for over 1 million deaths each year. As a bacterial pathogen, *Mtb* is highly adapted to survival within the host environment, which imposes numerous environmental stresses on the bacteria. To survive these challenges, *Mtb* utilizes a network of transcription factors to adapt its gene expression and physiology. CarD is an essential transcription factor in mycobacteria that binds directly to RNA polymerase (RNAP) rather than DNA sequence motifs on the chromosome.

Prior mechanistic studies of CarD's function in mycobacteria have been limited to studies of CarD's effects on individual kinetic rate constants in a handful of promoter contexts. However, chromatin-immunoprecipitation sequencing (ChIP-seq) studies indicate that CarD is present broadly across the mycobacterial genome. Therefore, to gain a holistic understanding of CarD's

role at all mycobacterial promoters, we performed RNA-sequencing (RNA-seq) on a set of *Mtb* and *Mycobacterium smegmatis* strains with point mutations causing defects in CarD's function. Our study demonstrated that over 50% of the *Mtb* genome is transcriptionally responsive to CarD mutation and revealed a novel bi-functional role of CarD as both a transcriptional activator and a transcriptional repressor. I performed bioinformatic analyses to show that CarD-regulated promoters could not be defined by a specific DNA sequence motif but that the direction of regulation was correlated with promoter characteristics such as -10 element conservation and discriminator G+C%. Our RNA-seq dataset raised a new question – 'How does CarD achieve regulatory specificity without binding specificity?'. In collaboration with the Galburt Lab, we developed a model in which the direction of CarD regulation is dictated by a promoter's basal transcription initiation kinetics rather than transcription factor binding. Specifically, our model predicts that CarD would activate transcription from promoters that form an unstable promoter open complex (RP_o) and repress transcription from promoters that form a stable RP_o by inhibiting the step of promoter escape.

To test this model, I performed *in vitro* transcription studies to determine CarD's direct regulatory effect on a panel of promoters with varying levels of basal RP_o stability. I showed that, indeed, while CarD activated transcription from the wild-type ribosomal RNA promoter *Mtb rrnA*P3, CarD repressed transcription from more stable promoter contexts. Furthermore, I showed that RP_o stability and CarD regulation could be influenced by sequence-independent factors such as DNA supercoiling, revealing another dynamic of transcription regulation.

Using the wealth of transcriptomic data that I gathered, I examined the overlap between CarD's *in vivo* regulon with other known regulons to gain insight into CarD's role in mycobacterial

physiology. Although I could not identify any specific downstream transcriptional pathway that was uniformly differentially expressed, CarD's regulon overlapped with numerous stress-responsive regulons including the DosRST two-component system and the alternative sigma factors σ^F, σ^H, and σ^M.

My thesis work advances our understanding of how RNAP-binding bacterial transcription factors like CarD are capable of enacting specific transcriptional responses over a broad regulon. My research also highlights how a single transcription factor can exert multiple functions under different contexts including contexts that cannot be understood from DNA sequence alone. Lastly, my work on CarD may help inform future design of *Mtb* therapeutics by deepening our understanding of the transcriptional programs that *Mtb* uses to endure environmental stresses.

Chapter 1: Introduction

Dennis X. Zhu, Christina L. Stallings

History and Pathogenesis of *Mycobacterium tuberculosis*

Tuberculosis is an ancient disease that has progressed with humans throughout the history of medical science. Some of the earliest evidence of tuberculosis in humans comes from Egyptian mummies with signs of Pott's disease, which is a disseminated infection of tuberculosis to the spine (1). Since then, tuberculosis, which has also been known by other names such as "phtisis" or "consumption", has been a focus of research for medical science from ancient Greece to modern times (2). A milestone for medical microbiology occurred in the late 19th century when German physician Robert Koch identified *Mycobacterium tuberculosis* (*Mtb*) as the etiological agent of tuberculosis (3), leading to the establishment of Koch's postulates. Despite its long relationship with human medicine, tuberculosis remains to be a threat to human health. According to WHO estimates, tuberculosis was responsible for roughly 1.3 million deaths worldwide in 2020 (4). Although there has been progress in combatting tuberculosis, and global incidence rates have generally declined over the last few decades (4), these advances in eradicating tuberculosis have been challenged by multiple factors including the emergence of drug-resistant strains of *Mtb*, concurrent health crises such as the COVID-19 pandemic, and the increase in the prevalence of comorbidities such as HIV, type 2 diabetes, and obesity.

Mtb is a rod-shaped, acid-fast bacteria in the phylum Actinobacteria and is the eponymous member of a group of genetically related species known as the *Mycobacterium tuberculosis* complex (MTBC) that are responsible for the human disease tuberculosis. While most species of bacteria in the genus *Mycobacterium* are environmental bacteria, the *Mycobacterium* genus contains the human pathogens *Mtb*, *M. leprae*, and *M. ulcerans* as well as several opportunistic pathogens that can causes disease in immunocompromised humans. Unlike other MTBC species, *Mtb* is an obligate pathogen of humans with no known animal reservoirs, and the evolutionary

history of its genome reflects its adaptation towards humans (5, 6). Compared to the most closely related non-tuberculous mycobacteria (NTM) species *M. kansasii*, which rarely infects humans, *Mtb* has a much smaller genome of 4.4Mb compared to 6.4Mb in *M. kansasii*, indicative of a genomic downsizing as *Mtb* evolved as a specialized human pathogen (5, 6). During its vertical evolution from the MTBC ancestor, *Mtb* also acquired a number of genes by horizontal gene transfer (HGT) involved in modification of lipids and metabolic survival during anaerobic conditions (7). However many genes thought to be involved in host survival, including the PhoPR and DosRST two-component systems, mce-associated genes, and the type VII secretion system are conserved in NTM species, suggesting these virulence pathways have been co-opted for MTBC growth in humans (5). MTBC genomes also encode a disproportionately high number of toxin-antitoxin systems, which have been implicated in adaptation to host stresses (8). Overall, the *Mtb* is a bacterial pathogen that is highly genetically adapted for its life within human hosts.

Tuberculosis infection in humans typically begins with the transmission of *Mtb* bacteria through aerosol droplets that are expelled by the cough of an infected individual with active disease and that are then inhaled by a susceptible host. Once inhaled the bacteria can establish a primary infection in the lung by trafficking to the interstitial space either by either invading the alveolar epithelium or via infection of a trafficking alveolar macrophage (9–11). Immediately, the bacterial cell surface is challenged by surfactant and antimicrobial peptides present in the mucosal surface of the lung epithelium (12). Although it is presumed the complex mycobacterial cell envelope plays a role in resisting cell surface stress, the potential role of active bacterial responses to cell surface stress are not well understood (13). Once the bacteria have been phagocytosed, the host cell attempts to degrade its bacterial cargo by acidification and production of reactive oxygen and nitrogen species (ROS and RNS). However, *Mtb* secretes numerous virulence factors that arrest

the phagosome-lysosome fusion pathway within the cell and block the recruitment of v-ATPase and NADPH oxidase to the phagosome, which induce low pH and ROS conditions, respectively (12–14). Furthermore, wild-type *Mtb* are naturally resistant to killing by low pH (15) and encode a number of factors including NuoG, SodA, and KatG that can detoxify ROS and RNS (13, 14, 16).

After failing to kill *Mtb* at a cell intrinsic level, the infected macrophage initiates a persistent inflammatory cascade that triggers the formation of a structure called a granuloma (17). A typical tuberculosis granuloma is comprised of a central necrotic core, called a caseum, surrounded by layers of differentiated macrophages with an epithelioid phenotype. The spatial landscape of the granuloma is marked by a gradient of inflammatory lipids, with the center containing pro-inflammatory eicosanoids and the periphery containing anti-inflammatory, tissue-healing lipid and transcriptional signals (18). Granulomas within the lungs of a single infected individual also display a range of immune control outcomes, with some lesions reaching sterilizing bacterial clearance while other lesions cannot control bacterial replication and eventually collapse to release *Mtb* bacilli capable of disseminating (19, 20). Historically, this "tug-of-war" between *Mtb* replication and immune control within the granuloma is thought to produce a stalemate that manifests as latent tuberculosis infection, where the bacteria are replicatively dormant but capable of reactivating many years after a primary infection (21). However, a recent meta-analysis of epidemiological data shows that 97% of active tuberculosis cases occur within two years of primary infection and the existence of an extended phase of *Mtb* dormancy is unlikely (22). Thus, the initial events between *Mtb* bacilli and host immunity during primary *Mtb* infection represent the critical juncture of disease outcome. Our understanding of how *Mtb* responds to this initial

onslaught of host derived stresses and heterogenous environments is key to advancing treatments for tuberculosis.

Transcription Regulation in Bacteria

Structure of the Bacterial RNA Polymerase

Gene expression is the ability of a living organism to produce functional proteins or RNAs from the genes that are encoded in DNA. Transcription of all genes in bacteria is performed by a single DNA-dependent RNA polymerase (RNAP) that consists of five core protein subunits – β, β', two α subunits, and ω. The binding of an additional, dissociable σ-subunit to the core enzyme forms the RNAP holoenzyme and allows for DNA sequence recognition at promoter regions. Our understanding of the structure of this complex has primarily been established through crystal structures of the *Thermus thermophilus* RNAP (23–25), but structures of RNAP from other species including mycobacteria (26–29) have become increasingly available in the last few years. The overall structure of the core enzyme resembles a "crab-claw" with the β and β' subunits forming two sides of the claw that is joined at its base by an α-subunit homodimer (α_2). The N-terminal domains (NTDs) of α_2 bind to β and β' and serve as a scaffold for RNAP assembly. The ω subunit is the smallest subunit of the core enzyme and dispensable for viability *in vivo* and for transcription *in vitro*. Although the role of the ω subunit remains ambiguous in many bacteria, the *Mtb* ω subunit contains a non-conserved loop that is essential for core enzyme assembly (30). Once assembled, the RNAP core enzyme creates a flexible protein complex with several openings that allow for the entry and exit of substrates during transcription. The large, positively charged cleft between the β and β' subunit "claws" is known as the "primary channel" and houses the DNA template and the nascent RNA-DNA hybrid. As DNA winds through the primary channel, it approaches the active

site of the polymerase at the back wall of the cleft near the hinge of the crab claw-like structure. The active site is defined by a catalytic Mg^{2+} ion that is coordinated by a triad of aspartic acid residues on the β' subunit. Near this site is a smaller opening known as the "secondary channel", which allows for the entry of NTP substrates to the active site. The secondary channel also serves as the binding site for numerous regulatory molecules including DksA, Gre factors, and some small molecule inhibitors of transcription (31). Lastly, the "RNA exit channel" is also located at the back of the primary cleft and serves to separate the RNA-DNA hybrid during transcription. The RNA exit channel also regulates transcription initiation, pausing, and termination through interactions with the secondary structure of newly synthesized RNA and protein factors such as Rho (32, 33)

While the core RNAP is capable of transcribing RNA, promoter-specific transcription initiation requires the binding of a σ factor. Our understanding of RNAP holoenzyme structure and function is primarily derived from complexes with the primary *E. coli* housekeeping σ factor σ^{70} and *E. coli* (σ^{A} in most other species). The σ^{70} family of proteins can be classified based on the presence of four structural regions: $\sigma_{1.1}$, σ_2, σ_3, and σ_4. Group 1 σ factors contain all four regions and comprise the housekeeping σ factors encoded in all bacterial genomes. Group 2 σ factors lack region 1.1 and are typically associated with stress responsive transcription but can also transcribe housekeeping genes (34, 35). Group 3 and 4 σ factors typically transcribe small regulons of genes dedicated to specific environmental situations. Group 4 σ factors are also known as extracytoplasmic function (ECF) σ factors and lack regions 1.1 and 3. ECF factors are the most numerous and functionally diverse group of σ factors. Each of the σ factor structural regions plays a different role in holoenzyme function. Overall, σ^{70} binds on the on the outside surface of the β' subunit with only region $\sigma_{3.2}$ winding through the primary channel past the active site. The σ factor is oriented with the N-terminus facing downstream DNA and the opening of the primary channel

and the C-terminus facing upstream DNA and the α_2 "hinge" of the crab claw. While σ is unbound to the core enzyme, the flexible region $\sigma_{1.1}$ blocks other regions of the protein from binding DNA in the absence of core RNAP (36). Once bound, $\sigma_{1.1}$ blocks the active site of the primary channel until it is dislodged by conformational changes from σ-DNA interactions from other regions (37), thus preventing non-specific initiation by RNAP holoenzyme. Regions σ_2, σ_3, and σ_4 face outwards from β' and are positioned to interact with DNA sequences in the -10 element, extended -10 element, and -35 element of promoter DNA, respectively (23). These interactions form the basis of initial promoter recognition by RNAP holoenzyme (32, 34).

The Process of Transcription Initiation

Transcription initiation is the first step of gene expression, and therefore the rate of this process is a key determinant of RNA production within bacteria and a focus of extensive study (38–40). The process begins with the binding of RNAP holoenzyme to a region of promoter DNA to form a closed promoter complex (RP_c). Since the DNA is in a duplex state at this point, the initial interactions are weak, but DNA footprinting studies suggest that multiple interactions are progressively established from upstream DNA towards the transcription start site (TSS) (41). The furthest upstream interaction occurs between the α-subunit C-terminal domains (αCTDs) and an A-T-rich sequence found in some promoters known as the UP element (42). This initial binding allows for further interactions between RNAP and the -35 element and the -10 element to form. These interactions trigger a series of conformational changes, collectively called "isomerization", in both RNAP and the DNA molecule that result in the bending of promoter DNA near the -35 element to enter the RNAP cleft and melting of a 12-15 base pair transcription bubble (38, 39). During isomerization, the RNAP β and β' subunits act as a dynamic clamp that opens and closes to first accommodate the DNA template into the RNAP cleft and then to progressively nucleate

and expand the transcription bubble (43, 44). Eventually, the complex forms a promoter open complex (RP_o) where the template strand +1 TSS base is placed in the active site and the bubble is stabilized by a closed clamp (45). A combination of kinetic (28, 46) and cryo-electron microscopy (cryo-EM) (26) studies of the *Mtb* RNAP indicate that the process of transcription bubble melting occurs in at least two steps through a partially melted intermediate with the transcription bubble opened from -11 to -4 (RP_i). Once the +1 TSS is loaded into the active site, RNAP begins RNA synthesis to form a series of ternary complexes in which RNAP holoenzyme is stably bound at the promoter with a nascent RNA-DNA hybrid in the RNAP primary channel (collectively termed RP_{itc}). As RNAP undergoes *de novo* RNA synthesis, downstream DNA is pulled into the RNAP cleft as the transcription bubble expands in a process known as "scrunching" (47). Once the RNA transcript reaches a critical length, it clashes with $\sigma_{3.2}$, which normally occludes the RNA exit channel, and disrupts contacts between RNAP and promoter DNA leading to promoter escape and ejection of the σ subunit from the RNAP elongation complex (48). Current kinetic data supports a branched model of transcription initiation (49, 50), where some initiation complexes never undergo promoter escape, instead cycling between RP_{itc} and RP_o while synthesizing and releasing a short RNA product in a process called abortive initiation (47).

The complicated multi-step process of transcription initiation is driven forward by binding free energy and a network of dynamic DNA-protein interactions that evolve throughout the pathway (32). Thus, the quantity and quality of these interactions at a given promoter, determined by DNA sequence, can greatly influence the kinetics of transcription initiation (34). As the transcription bubble starts to melt during isomerization, single-stranded DNA (ssDNA) residues are exposed to interact with different regions of RNAP in a sequence-specific manner. Many of these interactions occur during transient intermediates, so our understanding of the critical

interactions come from crystal structures of RNAP-σ^{70}/RNAP- σ^{A} RP_o (27, 45) and melting intermediates that have been captured by cryo-EM (26, 51). Bubble nucleation is driven by a sequence specific interaction between $\sigma_{2.3}$ and an A_{-11} on the non-template strand that is conserved in σ^{70} -10 elements (consensus $T_{-12}ATAAT_{-7}$) (52). Two highly conserved residues, A_{-11} and T_{-7}, are flipped out into binding pockets in $\sigma_{2.3}$ and are critical for bubble formation (53). These initial interactions in the -10 contribute to a 90° bend of downstream DNA that positions it in the active site (53). In RP_o, $T{:}A_{-12}$ remains un-melted and is bracketed by a pair of tryptophan residues in $\sigma_{2.3}$, representing the upstream edge of the bubble (27, 45). Interactions between σ_4 and the conserved TTG motif in the -35 element (consensus $T_{-35}TGACA_{-30}$) are maintained in RP_o (45, 54). An less conserved region known as the extended -10 (consensus $T_{-15}G_{-14}$) can form a sequence-specific interaction with $\sigma_{3.2}$ (45) and has been implicated in transcription strength at promoters that lack a consensus -35 element (55). Further downstream, in the discriminator region, which is the region between the -10 element and the TSS, $\sigma_{1.2}$ can form sequence-specific interactions with a $G_{-5}GA_{-3}$ motif (56, 57). At the downstream edge of the transcription bubble, the core RNAP makes contacts with multiple bases from -4 to +2, which is a region known as the core recognition element. Specifically, RNAP β forms a pocket that can bind a melted G_{+2} at the upstream boundary of the bubble (56). Collectively, the presence or absence of these interactions can affect multiple kinetic rates during transcription initiation, highlighting how DNA sequence context contributes to the wide variance in transcription initiation rates at different promoters (38).

Unique aspects of the mycobacterial transcription machinery

While our understanding of transcription initiation is founded in studies of the *E. coli* σ^{70} holoenzyme, many bacterial species have lineage-specific characteristics in their promoter architecture or RNAP structure that diverge from the paradigms established in *E. coli* (40, 58).

Compared to the *E. coli* RNAP, mycobacterial RNAP form relatively unstable RP_o in the same promoter contexts (46, 59). Multiple differences between the bacteria could potentially explain this divergence in transcription initiation kinetics including lineage-specific structural insertions in RNAP, differences in promoter architecture, and the presence of additional molecular factors. While the overall structure of RNAP core enzyme is highly conserved across bacteria, the mycobacterial RNAP contains a lineage-specific β'i1 insertion consisting of two anti-parallel α-helices that extend from the tip of the RNAP clamp (27, 29). The group 1 σ factor of *Mtb* σ^A also contains a long N-terminal extension in region $\sigma_{1.1}$ that is over twice the length of *E. coli* $\sigma^{70}_{1.1}$ (27). Structures of *Mtb* RP_o and RP_{itc} show interactions between β'i1, $\sigma^A_{1.1}$, and downstream double-stranded DNA (dsDNA) near the entrance of the active site cleft, and deletion of either β'i1 or $\sigma^A_{1.1}$ greatly decrease RP_o stability (29). Together these data suggest that the mycobacteria-specific inserts in RNAP β' and σ^A act as an additional "gate" to help stabilize the RNAP clamp module in RP_o. In terms of promoter architecture, the -35 element 'TTG' motif is poorly conserved in Actinobacteria (58, 60). However, mycobacterial promoters are enriched for UP elements and other lineage-specific motifs such as a G at position -13 that may compensate for the lack of conservation in the -35 element (27, 61). Lastly, in each species, its transcription machinery operates within a unique environment of species-specific regulatory proteins and nucleic acids. CarD and RbpA are two essential proteins found in mycobacteria but absent from *E. coli* that bind to RNAP during transcription initiation and are required for RP_o stabilization (28, 62–64). These inter-species differences highlight the necessity of studying transcription mechanisms within their native contexts.

Factors that regulate transcription initiation

Over the lifetime of a bacterium, the cell encounters numerous environmental stimuli that require a physiological response. Most often in bacteria, these responses are regulated at the level of transcription initiation. The most direct mechanism by which bacteria alter their gene expression is through the usage of alternative sigma factors. *Mtb* encodes 13 sigma factors including the housekeeping group 1 sigma factor σ^A, the group 2 sigma factor σ^B, and eleven ECF sigma factors $\sigma^{C\text{-}M}$ (65, 66). Each alternative sigma factor recognizes a different -10 and -35 element sequence motif, and their binding to RNAP core alters its promoter specificity. σ^A is the only essential sigma factor in *Mtb* and is most abundant sigma factor during exponential growth. σ^B is dispensable for growth and up-regulated during stationary phase, leading the field to characterize its role as a stress-responsive transcription factor. However, chromatin-immunoprecipitation sequencing (ChIP-seq) studies indicate that σ^A and σ^B localize to an overlapping set of promoters, suggesting that σ^B may also function to transcribe housekeeping genes albeit under different regulatory conditions than σ^A (35). The group 4 ECF sigma factors of *Mtb* are generally function during specific stress conditions and their activity is often regulated post-translationally by anti-sigma factors, which bind and sequester their cognate sigma factor (67, 68). In addition to anti-sigma factors, there is a growing appreciation for a class of σ-subunit remodeling factors that bind to sigma factors to alter its assembly with or activity on RNAP core (69). RbpA is one such factor that is unique to *Actinobacteria* and essential in *Mtb* (62, 70). Structural studies show that RbpA promotes the assembly of RNAP-σ^A/σ^B holoenzymes by binding to region 2 of the sigma factors to stabilize them in an active conformation that is capable of binding RNAP core and promoter DNA (28, 71–73). In addition to its sigma factor remodeling function, RbpA also makes numerous additional contacts with RNAP and DNA to stabilize RP_o (64, 74), alter transcription initiation

kinetics (28, 75), and create a binding pocket for the transcription-targeting antibiotic fidaxomicin (76, 77).

In addition to alternative sigma factors, bacteria also utilize a repertoire of transcription factors that bind to specific promoters to alter the activity of RNAP, herein referred to as classical transcription factors (78, 79). *Mtb* encodes over 150 DNA-binding transcription factors (80, 81). Many classical transcription factors act by altering the association of RNAP holoenzyme with specific promoters, either by steric hinderance (classical repressors), binding near UP elements and interacting with RNAP-αCTD (class I activators), or binding near the -35 element and interacting with σ_4 (class II activators) (79). Some factors, such as Fis and IHF in *E. coli*, blend the lines between nucleoid associated proteins (NAPs) and transcription factors (82). While these factors do not interact with RNAP, they recognize DNA in a sequence-specific manner and can alter local DNA topology to induce or inhibit RNAP-promoter interactions. A key feature of transcription factors is their ability to respond to environmental cues. Some factors can sense directly sense their environmental stimulus. For example, the WhiB-like family of transcription factors in *Actinobacteria* directly sense the redox state of the cell through an iron-sulfur cluster and transduce a transcriptional response by binding to RNAP and DNA to stabilize RP_c (83, 84). In contrast, transcription factors belonging to two-component systems are coupled to a separate sensor histidine kinase protein that transduces the information from environmental cues to a response regulator, often through a phosphorylation cascade. *Mtb* has twelve complete two-component system that respond to a range of cues including hypoxia, pH, and phosphate starvation (85). The regulons of classical transcription factors are defined by the presence of a DNA-sequence motif that allows for transcription factor binding.

Classical DNA-binding transcription factors are complemented by a growing class of molecular factors that bind directly to RNAP to modulate its activity, herein referred to as RNAP-binding transcription factors (86). Since these factors do not require a specific DNA sequence motif, they often are able to regulate a broader range of promoters than classical transcription factors. Some of these factors act as monotonic activators or repressors such as the 6S RNA, which is a small regulatory RNA that mimics DNA in the transcription bubble and sequesters σ^{70} holoenzymes during stationary phase in *E. coli* to down-regulate housekeeping transcription (87). However, certain RNAP-binding transcription factors are capable of potentiating more complex patterns of promoter-specific up- and down-regulation of transcription. The most well-characterized example is the *E. coli* stringent response factors DksA and guanine (penta)tetraphosphate [(p)ppGpp] (88, 89). In *E. coli*, (p)ppGpp is synthesized in response to nutrient starvation and binds along with DksA to RNAP to directly inhibit transcription from promoters of stable RNAs and activate transcription of amino acid biosynthesis operons (89–91). DksA/(p)ppGpp accomplish this by de-stabilizing an isomerization intermediate prior to stable RP_o formation, thus repressing transcription from promoters that struggle to maintain stable RP_o while activating transcription from promoters with stable RP_o by increasing the forward rate of isomerization (92–95). This example demonstrates how RNAP-binding transcription factors can discriminate promoters by their underlying kinetics to potentiate multiple regulatory outcomes.

CarD: A Conserved RNAP-Binding Transcription Factor

Molecular description of CarD

CarD is a small, 162 amino acid RNAP-binding transcription factor that is essential for viability in *Mtb* and required by the mycobacterial transcription machinery to form stable RP_o at

its ribosomal RNA (rRNA) promoters (46, 59). Structurally, CarD is a globular protein with two lobes representing its two functional domains, a β-stranded N-terminal RNAP interaction domain (RID) and an α-helical C-terminal DNA-binding domain (DBD) (25, 96). CarD's RID adopts a conserved fold resembling the RID of TCRF and interacts directly with the surface of RNAP β subunit β1-lobe (97–99). The DBD of CarD does not resemble the conserved helix-turn-helix DBD fold found in other bacterial transcription factors. Instead, CarD's DBD interacts with DNA through a positively charged patch of basic amino acids contributed by three distinct α-helices (98). Mutations to CarD that disrupt this charged patch also disrupt its interaction with DNA (98, 100). In structures of CarD bound to *Mtb*RNAP in RP_o and a partially melted intermediate, CarD's DBD is positioned to interact with DNA at the dsDNA/ssDNA junction at the upstream edge of the transcription bubble (26). At this junction, a conserved tryptophan residue (W85) wedges into a the minor groove of the DNA backbone and interacts with T_{-12} on the non-template strand to stabilize RP_o (25, 96, 100). Although there are no published structures of CarD bound to *Mtb*RNAP in the closed complex, structural predictions suggest that CarD-W85 would sterically clash with dsDNA, leading some to hypothesize that this clash underlies CarD's positive effect on the forward rate of isomerization (25, 46). Based on DNA footprinting and comparative structural analysis of RP_o with and without CarD bound, CarD binding does not appear to change the overall structure of the mycobacterial transcription bubble (45, 59, 96). Thus, CarD's effects on RP_o stability are structurally explained by the network of interactions between CarD-RID and RNAP-β1, CarD-DBD and DNA, and the positioning of CarD-W85 at the edge of the transcription bubble.

CarD homologs in other bacterial species

In *Mtb*, mutation or depletion of CarD results in pleiotropic phenotypes that range from sensitivity to oxidative stress, antibiotic stress, nutrient starvation, DNA-damage, and reduced

survival during mouse infection (63, 97, 100–102). A ChIP-seq study of CarD binding in *M. smegmatis* indicates that it is broadly co-localizes with RNAP-σ^A to numerous promoter regions (25, 103), in agreement with its ubiquitous effect on bacterial physiology. While CarD functions as a broad, essential transcription regulator in mycobacteria, it occasionally plays more specialized roles in other bacteria. Homologs of CarD are widely distributed across bacterial phyla but absent from *E. coli* (63) (**Table 1**). Outside of mycobacteria, CarD was first characterized in *Myxococcus xanthus* as CdnL (CarD N-terminal-like), named for its similarity to the N-terminal to a separate *M. xanthus* protein named CarD (104). Unlike *M. xanthus* CarD, which is a non-essential transcription regulator involved in carotenogenesis, CdnL is essential in *M. xanthus* and appears to be involved in broad homeostatic processes (104, 105). In the α-proteobacterium *Caulobacter crescentus*, the CarD homolog *Ccr*CdnL is non-essential, but Δ*cdnL* mutants display severe phenotypes in cell morphology, central carbon metabolism, and cell wall biosynthesis, indicating that *Ccr*CdnL is still critical to cellular homeostasis even if is it non-essential (106, 107). In other bacteria such as *Bacillus cereus* (108) or *Borrelia burgdorferi* (109), the CarD homolog is only required during a specific phase of the bacterial life cycle. Although CarD's role in physiology remains relatively vaguely defined across different bacteria, CarD's role in regulating rRNA biosynthesis appears to be a conserved function (63, 107, 109, 110).

Studies of CarD's kinetic mechanism in mycobacteria

CarD's effects on the kinetics of transcription initiation in mycobacteria have been illuminated by numerous *in vitro* studies over the past decade (**Table 2**). Early studies established several key points about CarD's *in vitro* activity: 1. CarD stabilizes RP_o formed by mycobacterial RNAP (25, 59, 96, 100, 101), 2. CarD inhibits bubble collapse (46, 59, 96), 3. CarD promotes bubble nucleation (28, 46). β-galactosidase promoter activity assays in *M. smegmatis* show that

CarD activates transcription from mycobacterial rRNA promoters (25, 100). Recently, CarD has also been shown to inhibit the rate of promoter escape by *Mtb*RNAP (75, 111). Detailed analysis of a transient fluorescence peak in the promoter escape reaction of *Mtb*RNAP suggests that RbpA inhibits the initial synthesis of a short pre-escape RNA product while CarD inhibits the final step of promoter clearance (75). These divergent roles of CarD and RbpA during promoter escape match their structural position on RNAP; while RbpA threads through the primary channel near the active site (28), CarD is bound to the outer surface of RNAP-β distant from the active site and not contacting σ (25). Collectively, these observations have led to the hypothesis that CarD generally functions as a transcriptional activator, but this has yet to be directly tested *in vitro*. Furthermore, *in vitro* studies of CarD's kinetic mechanism have been limited to a handful of promoters, leaving a gap in our understanding of how CarD functions in other promoter contexts.

Tables

Organism	Gene name	Essentiality	Studies	Hypothesized functions
Mycobacterium tuberculosis	*carD*	Essential	Stallings *et al*. 2009[1]; Weiss *et al*. 2012[2], Garner *et al*. 2014[3], Garner *et al*. 2017[4], Zhu *et al*. 2019[5], Li *et al*. 20226	General growth[134], stringent response[16], oxidative stress[1236], nutrient starvation[15], DNA damage[16], antibiotic tolerance[1236], virulence in mice[1234], ribosomal RNA synthesis[134], transcriptional homeostasis[5]
Caulobacter crescentus	*cdnL*	Non-essential	Gallego-García *et al*. 2017[1], Woldemeskel *et al*. 2020[2]	General growth[12], cell morphology[12], ribosomal RNA synthesis[12], central metabolism[2], cell wall synthesis[2]
Myxococcus xanthus	*cdnL*	Essential	García-Moreno *et al*. 2010[1], Gallego-García *et al*. 2014[2], Bernal-Bernal *et al*. 2015[3]	Cell morphology[1], cell division[1], transcription of housekeeping genes[2], activating ECF-dependent transcription[3]
Rhodobacter sphaeroides	*carD*	Essential	Henry *et al*. 2020[1], Henry *et al*. 2021[2]	Ribosomal RNA synthesis[1], general growth[1], auto-regulation[2]
Borrelia burgdorferi	*ltpA*	Non-essential	Chen *et al*. 2018	Growth in ticks, ribosomal RNA synthesis, virulence in mice
Liberibacter sp.	*prbP*	Unknown	Gardner *et al*. 2016[1], Pan *et al*. 2021[3]	Virulence in plants[1], cell division[2], cell motility[2], biofilm formation[2]
Bacillus cereus	*cdnL*	Non-essential	Warda *et al*. 2016	Recovery of heat damaged spores

Table 1. Studies of CarD homologs in different bacterial species.

Study	Assays Used	Promoters Used	Conclusions
Srivastava *et al*. 2013 (25)	β-galactosidase assay in *M. smegmatis*[1] Single-round *in vitro* transcription[2] Promoter binding assay[3]	*M. smegmatis rrnA*P123[1] *Mtb rrnA*-AP3[23] *T. thermophilus* 23S promoter[23]	CarD functions as a transcription activator *in vivo* CarD stimulates promoter binding
Garner *et al*. 2014 (100)	β-galactosidase assay in *M. smegmatis*[1] Complex stability assay[2]	*M. smegmatis rrnA*P1[1], P2[1], P3[1], and P123[2] *Mtb rrnA*P3[2] and P13[2]	CarD stabilizes competitor-resistant RNAP-promoter complex via RID and DBD
Davis *et al*. 2015 (59)	Single-round three-nucleotide transcription assay[1] Complex stability assay[2]	*Mtb rrnA*P3[1] *Mtb rrnA*P3 Δ23[1] AC50[12] AC50 synthetic bubble[2]	CarD stabilizes weak open complexes formed by mycobacterial RNAP relative to *E. coli* RNAP by preventing bubble collapse
Rammohan *et al*. 2015 (46)	Stopped-flow fluorescence open complex formation[1]	*Mtb rrnA*P3[1]	CarD stimulates open complex formation by promoting bubble opening and inhibiting collapse
Bae *et al*. 2015 (96)	Three-nucleotide transcription assay[1] Complex stability assay[2]	*T. thermophilus* 23S promoter[123] *Mtb rrnA*P3[12] *T. thermophilus* 23S promoter with synthetic bubble[3]	CarD increases open complex lifetime by preventing bubble collapse via conserved tryptophan residue (W85)
Rammohan *et al*. 2016 (64)	Stopped-flow fluorescence open complex formation[1]	*Mtb rrnA*P3[1]	CarD and RbpA cooperatively increase the speed and stability of open complex formation
Garner *et al*. 2017 (101)	Stopped-flow fluorescence open complex formation[1] Three-nucleotide transcription assay[2]	*Mtb rrnA*P3[12]	CarD stabilizes *M. bovis* RNAP open complex and increases formation of initial RNA products dependent on RNAP affinity
Hubin *et al*. 2017 (28)	Complex stability assay[1] Stopped-flow fluorescence open complex formation[2]	*Mtb rrnA*P3[12] *Mtb vapB* promoter[2]	CarD and RbpA stabilize open complex via unique mechanisms. CarD increases the rate of initial bubble nucleation and inhibits collapse.
Zhu *et al*. 2019 (111)	Promoter escape assay[1]	*Mtb rrnA*P3[1]	CarD slows the rate of promoter escape
Jensen *et al*. 2019 (75)	Stopped-flow fluorescence open complex formation[1] Stopped-flow fluorescence promoter escape[2]	*Mtb rrnA*P3[12]	CarD inhibits promoter escape from RP_{itc} while RbpA inhibits RP_{itc} formation

Table 2. Kinetic studies of CarD's activity on the mycobacterial transcription machinery.

Chapter 2: CarD contributes to diverse gene expression outcomes throughout the genome of *Mycobacterium tuberculosis*

Dennis X. Zhu, Ashley L. Garner, Eric A. Galburt, and Christina L. Stallings

A version of this chapter was originally published as (111).

Abstract

The ability to regulate gene expression through transcription initiation underlies the adaptability and survival of all bacteria. Recent work has revealed that the transcription machinery in many bacteria diverges from the paradigm that has been established in *Escherichia coli*. *Mycobacterium tuberculosis* (*Mtb*) encodes the RNA polymerase (RNAP) binding protein CarD, which is absent in *E. coli* but required to form stable RNAP-promoter open complexes (RP_o) and essential for viability in *Mtb*. The stabilization of RP_o by CarD has been proposed to result in activation of gene expression, however, CarD has only been examined on limited promoters that do not represent the typical promoter structure in *Mtb*. In this study we investigate the outcome of CarD activity on gene expression from *Mtb* promoters genome-wide by performing RNA-sequencing on a panel of mutants that differentially affect CarD's ability to stabilize RP_o. In all CarD mutants, the majority of *Mtb* protein encoding transcripts were differentially expressed, demonstrating that CarD had a global impact on gene expression. Contrary to the expected role of CarD as a transcriptional activator, mutation of CarD led to both up- and down-regulation of gene expression, suggesting that CarD can also act as a transcriptional repressor. Furthermore, we present evidence that stabilization of RP_o by CarD could lead to transcriptional repression by inhibiting promoter escape, and the outcome of CarD activity is dependent on the intrinsic kinetic properties of a given promoter region. Collectively, our data support CarD's genome-wide role of regulating diverse transcription outcomes.

Introduction

Bacterial pathogens must coordinate diverse transcriptional responses in order to survive the multitude of host-derived stresses that they encounter during infection. All bacteria encode a

single, multi-subunit RNA polymerase (RNAP) core enzyme that consists of the β, β', α_2, and ω subunits. Transcription initiation requires the binding of a dissociable σ factor to form the RNAP holoenzyme, which recognizes and binds promoter sequences upstream of the +1 transcription start site (TSS). Initially the holoenzyme binds to closed, double-stranded DNA (dsDNA) to form the RNAP-promoter closed complex (RP_c). The RP_c then undergoes a number of reversible conformational changes to form the RNAP-promoter open complex (RP_o), in which the promoter DNA is melted into a transcription bubble and the +1 TSS is positioned in the RNAP active site and accessible to the initiating nucleotide (iNTP). To complete transcription initiation and form a full-length RNA product, the RNAP must break interactions with DNA and escape the promoter. Our understanding of bacterial transcription initiation is largely formed by studies in *Escherichia coli*, but recent studies of these processes in other bacteria suggest that some paradigms do not hold true in all species. For example, *in vitro* kinetics analyses show that RNAP from *Bacillus subtilis*, *Thermus,* and *Mycobacterium* species form unstable RP_o compared to *E. coli* RNAP (46, 59, 112, 113). Furthermore, mycobacterial genomes possess divergent promoter architecture from *E. coli* promoters and are not well expressed in *E. coli* (60, 114). Thus, in order to obtain a holistic understanding of gene expression in bacteria, transcription initiation events should be studied in their native contexts.

Mycobacterium tuberculosis (Mtb) infection currently results in more deaths a year than any other single infectious agent (115). To develop effective therapeutics for *Mtb* infection and combat this epidemic, there is a need to better understand how *Mtb* regulates its gene expression to facilitate its survival in the host. Mycobacteria encode an essential RNAP-associated transcription factor CarD that is widely distributed across many eubacteria phyla but not conserved in *E. coli* (63, 68). In mycobacteria, *carD* expression is induced by genotoxic, oxidative, starvation,

and antibiotic stresses (63), suggesting that CarD may serve as a stress-responsive transcriptional regulator in addition to its essential function. CarD homologs in other bacteria have also been shown to respond to diverse environmental stimuli (109, 116) and regulate critical processes such as metabolic homeostasis and cell division (104, 107).

Chromatin immunoprecipitation-sequencing (ChIP-seq) experiments in mycobacteria revealed that CarD localizes with RNAP holoenzyme at promoters throughout the mycobacterial genome but not within gene coding sequences (CDS), indicating that CarD specifically functions during transcription initiation (25, 103). Depletion or mutation of CarD has been shown to lead to changes in ribosomal RNA (rRNA) levels (63, 97, 100) and, therefore, initial studies of CarD have focused on rRNA promoters. CarD stabilizes RP_o formed by mycobacterial RNAP-σ^A holoenzyme at the rRNA *rrnA*P3 promoter (25, 46, 59, 97, 100). CarD consists of two functional domains that are both required for stabilization of RNAP-σ^A-rrnAP3 RP_o (100). The N-terminus of CarD comprises an RNAP-interacting domain (RID) that interacts directly with the β1-lobe of the RNAP-β subunit (63, 97). CarD also possesses a C-terminal DNA-binding domain (DBD) containing a patch of basic residues that is predicted to interact with promoter DNA near the upstream edge of the transcription bubble (25, 96, 100). In addition, a conserved tryptophan (W85) within the C-terminal basic patch is also important for CarD's effects on RNAP-σ^A-rrnAP3 RP_o and, based on structural studies, has been proposed to wedge into the upstream edge of the transcription bubble to prevent bubble collapse (96, 100). *Mtb* strains encoding mutations in the CarD RID or DBD are attenuated in a mouse aerosol infection model and exhibit increased sensitivity to killing by oxidative stress and the antibiotics rifampin, streptomycin, and ciprofloxacin (97, 100, 101).

Investigations of CarD's activity thus far have been limited to a handful of promoters with architectures similar to promoters in *E. coli*, including *rrnA*P3. Therefore, the *in vivo* impact of CarD on mycobacterial gene expression genome-wide remains unknown. To gain a comprehensive view of how CarD regulates transcription globally at all *Mtb* promoters, we performed RNA-sequencing (RNA-seq) on strains of *Mtb* expressing mutants of CarD that either disrupt or enhance CarD activity. Our model of CarD activity based on studies of *rrnA*P3 predicts that CarD activates transcription by stabilizing RP_o. If this model held true for all *Mtb* promoters that CarD regulates, then we would expect that impairing CarD's activity would lead to global down-regulation of gene expression. On the contrary, we found that CarD mutants show an equal number of down-regulated transcripts and up-regulated transcripts, suggesting that CarD is capable of repressing transcription from certain *Mtb* promoters and that its role is more complex than that of a monotonic transcriptional activator. Our comparison of gene expression patterns shows that CarD mutants with similar effects on RP_o stability possess similar transcriptional profiles, indicating that the stabilization of RP_o by CarD is responsible for gene expression outcomes, which can be diverse and depend on the promoter. Our studies broaden our understanding of how CarD regulates transcription of numerous promoters *in vivo* to coordinate a gene expression profile that promotes bacterial viability.

Experimental Procedures

Bacterial strains and growth conditions. All *Mtb* strains were grown at 37°C in Sauton's broth media (0.5 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4$, 4g L^{-1} L-asparagine, 60 mL glycerol, 0.05 g L^{-1} ferric ammonium citrate, 2.0 g L^{-1} citric acid, 0.1 ml L^{-1} 1% $ZnSO_4$, 0.05% Tween 80, pH 7.0). All strains were derived from the Erdman strain and described in previous publications (97, 100, 101).

RNA preparation and sequencing library preparation. RNA was prepared from 30 mL of log-phase *Mtb* from $CarD^{WT}$, $CarD^{R47E}$, $CarD^{K125A}$, $CarD^{I27F}$, and $CarD^{I27W}$ in triplicate, as previously described (63). Each sample was treated with DNAseI using the TURBO DNA-free kit (Thermo-Fisher) and then 1-5µg of DNase-treated RNA was submitted to the Washington University Genome Technology Access Center (GTAC) for Illumina sequencing. Ribo-zero was used to deplete ribosomal RNA.

Read alignment and RNA-seq data analysis. Single-end RNA-seq reads were pseudoaligned to an index of *Mtb* H37Rv CDS (assembly ASM19595v2) and counted using *kallisto* (117). Data analysis was performed in R using the Rstudio environment. Un-normalized counts were imported using the *tximport* package and differential expression analysis was performed using the *DESeq2* package (118, 119). Differential gene expression was determined by a Benjamini-Hochberg adjusted p-value (p_{adj}) of <0.05. For visualization, $\log_2$ fold-changes were shrunken using the *apeglm* method (120). Hierarchical clustering in **Fig. 3** was performed based on Pearson correlation distance of $\log_2$ fold-change trends across four genotypes using Ward's method linkage. The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus (121).

RNA spike in controls and qRT-PCR. RNA was harvested from 40 mL of log-phase *Mtb* as previously described for RNA-seq. MS2 bacteriophage RNA (Roche) was added to each sample at a concentration of 1ng MS2 RNA per 1 billion bacterial cells. Each RNA sample was then treated with DNaseI (Thermo-Fisher Scientific), and cDNA was synthesized with SuperScript III First Strand Synthesis Kit (Invitrogen). qRT-PCR was performed using SYBR green qPCR kit (BioRad). Transcript abundance levels were measured for eight *Mtb* genes that were up-regulated

in $CarD^{R47E}$ or $CarD^{K125A}$ according to our RNA-seq analysis. Primers are listed in Table 1. Transcript levels were normalized to amount of MS2 transcript and compared to levels in $CarD^{WT}$.

Promoter Escape Assays. The *Mtb rrnA*P3 promoter (nucleotides 1,460,113-1,370,157 of the *Mtb* Erdman strain) was cloned into the pRLG770 plasmid using EcoRI and HindIII restriction enzyme sites. Plasmids were prepared from *E. coli* by midiprep (Qiagen) and phenol/chloroform extraction. Purified recombinant CarD protein was stored in a buffer of 20 mM Tris-Cl pH 8.0, 150mM NaCl, and 1mM BME. Promoter escape assays are based on those previously described (122, 123). Briefly, 20nM *Mycobacterium bovis* core RNAP was pre-incubated on ice with 40nM σ^A and 50ng of DNA template. The reaction was brought to 12.5uL by dilution so that the final solution contained 1x transcription buffer (10mM Tris-Cl pH 8.0, 10mM $MgCl_2$, 50μg mL^{-1} BSA, 40mM NaCl, and 1mM DTT). Transcription was initiated with simultaneous addition of NTPs (100μM GTP, CTP, and ATP; 10μM UTP; and 0.1μL [α-^{32}P]UTP) and 400nM of competitor DNA (double-stranded FullCon promoter DNA fragment). The NTPs initiate transcription from the RNAP-$\square^A$-*rrnAP3* complexes while the competitor prevents formation of new complexes, thereby allowing for only one round of transcription to occur. The reactions were stopped after designated times with 2x formamide buffer and run on a 6% urea-PAGE gel. Products were visualized by phosphorimagery and quantified (Image Gauge Program).

Results

Point mutations in CarD result in global changes in gene expression in mycobacteria. The impact of CarD activity on global gene expression in *Mtb* has yet to be investigated and previous microarray experiments performed in the nonpathogenic model organism *Mycobacterium smegmatis* during CarD depletion were confounded by cell death (63). Therefore, to measure the

effects of CarD activity on expression profiles in *Mtb*, we performed RNA-seq with RNA collected from *Mtb* strains that express wild-type (WT) CarD ($CarD^{WT}$), $CarD^{R47E}$ (a RID mutant with weakened affinity for RNAP (97)), $CarD^{I27F}$ or $CarD^{I27W}$ (RID mutants with increased affinity to RNAP (101)), or $CarD^{K125A}$ (mutation within the DBD predicted to weaken the affinity to DNA (100)) as the only *carD* allele (Fig. 1*A*). *In vitro*, mutations that weaken the CarD-RNAP or CarD-DNA interaction impair the ability of CarD to stabilize RNAP-$\square^{A}$-*rrnAP3* RP_o (46, 100) while mutations that increase the apparent affinity of CarD to RNAP potentiate its RP_o-stabilizing activity at lower concentrations of CarD (46, 101). For our RNA-seq analyses, RNA was collected from cultures growing exponentially in unstressed conditions where all of the *Mtb* strains are viable.

In all four *Mtb* strains with mutations in *carD*, more than half of the genome was significantly ($p_{adj} < 0.05$) differentially expressed in comparison to the $CarD^{WT}$ strain (Fig. 1*B*), consistent with CarD's global presence at mycobacterial promoters (25). Notably, in all four CarD mutants, there were a similar number of genes that were up-regulated as there were down-regulated (Fig. 1*B*). When we focused on the genes that were differentially expressed greater than 2-fold *(SI Appendix* Dataset S1), we found that in $CarD^{K125A}$ there were twice as many genes that were up-regulated (106) compared to down-regulated (58) while in $CarD^{I27W}$ the opposite trend was true where there were twice as many genes down-regulated (79) as there were up-regulated (37). In $CarD^{R47E}$ and $CarD^{I27F}$ the number of genes up-regulated >2-fold was similar to the number of genes down-regulated >2-fold. When we categorized differentially expressed genes based on Tuberculist functional annotations (124), they fell into diverse functional categories *(SI Appendix* Dataset S1), indicating that CarD activity affects multiple cellular processes.

Given that the ability of CarD to interact with the RNAP and DNA is required to stabilize the RNAP-σ^A-*rrnA*P3 RP_o *in vitro* (46, 100), one may expect that the RID and DBD work in tandem to stabilize RP_o and activate transcription at all promoters that CarD regulates. In this model, the R47E and K125A point mutations would cause a general decrease in transcript production while the I27F and I27W point mutations that enhance the ability of CarD to stabilize RP_o may cause a general increase in transcript production. However, this model is not supported by our RNA-seq data, which shows that all four mutants exhibit similar numbers of up-regulated and down-regulated transcripts (Fig. 1*B*). These data could mean that CarD is capable of promoting both activation and repression of gene expression *in vivo*. However, another possible explanation for the RNA-seq data is that mutation of CarD caused a global up- or down-regulation of transcript production that we were unable to detect due to the submission of equal amounts of RNA from each sample for the RNA-seq reactions. To explore this possibility, we performed spike-in control experiments (125) where we isolated RNA from log-phase cultures of *Mtb* expressing $CarD^{R47E}$, $CarD^{K125A}$ and $CarD^{WT}$ and added 1ng MS2 bacteriophage RNA (Roche) per 1 billion cells to the RNA samples to serve as a proxy for cell number. cDNA was generated for each sample and quantitative reverse transcription PCR (qRT-PCR) was performed to determine the transcript level per cell of 8 *Mtb* genes that were either up or down-regulated in the $CarD^{R47E}$ and $CarD^{K125A}$ RNA-seq samples. The qRT-PCR results largely recapitulated our RNA-seq results (*SI Appendix* Fig. S1*)*, supporting that transcripts can be up or down-regulated in the CarD mutants on a per cell basis. The fold-change differences detected by qRT-PCR had greater variance than those detected by RNA-seq, which may reflect the additional experimental variability introduced when estimating cell number and adding spike-in RNA transcript. The presence of up-regulated genes in the CarD

mutant strains provides the first experimental evidence that CarD may repress gene expression at specific promoters in the *Mtb* genome.

CarD delays promoter escape and could lead to repression of gene expression depending on the kinetics of transcription initiation. The ability of a single factor that affects RP_o stability to lead to both activation and repression of gene expression is not unprecedented. For example, ppGpp and the proteobacterial transcription factor DksA have been shown to activate transcription from amino acid synthesis genes and repress transcription from genes encoding stable RNAs in *E. coli* by altering the transition rates between RP_c and RP_o (92, 95, 126). Transcript flux modeling suggests that CarD could lead to repression of transcript production via its ability to stabilize RP_o if there was a resulting decrease in promoter escape (127), which is predicted to occur based on the known correlation between RP_o stability and promoter escape rate (128). To determine if CarD can lead to a delay in promoter escape, we monitored the rate of formation of full-length transcript by *Mycobacterium bovis* RNAP-σ^A holoenzyme from the *Mtb rrnA*P3 promoter using single round *in vitro* transcription assays. In these assays, we added dsDNA competitor and NTPs at the same time to preformed RNAP-σ^A-*rrnAP3* complexes in the presence or absence of CarD. The NTPs initiate transcription from the RNAP-σ^A-*rrnAP3* complexes while the competitor prevents formation of new complexes, thereby allowing for only one round of transcription to occur. To estimate the rate of promoter escape, the time required for 90% of the final amount of transcript to be formed was calculated (Fig. 2*A*). In these assays, $CarD^{WT}$ significantly delayed promoter escape from the *rrnA*P3 promoter from 8.52 minutes to 13.07 minutes (Fig. 2*A*), demonstrating that CarD activity leads to both an increase in RP_o stability and a decrease in the rate of promoter escape.

In the context of this model, whether transcription is activated or repressed by CarD depends on the basal kinetics of transcription initiation from that promoter. Using an expression

for the steady-state flux of transcript production (127), the experimentally measured rates of RP_o formation (46), and the rate of promoter escape determined above, we calculated the ratio of RNA production in the presence and absence of CarD across sets of promoters where RP_o stability varied over four orders of magnitude. More specifically, the rate constants leading out of RP_o (k_{close} and k_{escape}) were systematically titrated together such that low rates represent relatively stable RP_o and high rates represent relatively unstable RP_o. In the presence of CarD, the rate constant of promoter opening (k_{open}) was increased 25-fold and the rate of bubble collapse (k_{close}) was decreased 8-fold, based on our previously published kinetic model (46). Furthermore, the rate of escape (k_{escape}) was decreased 1.5-fold based on the ratio of escape rates determined experimentally from our gel-based assay (Fig. 2*A*-*B*).

The results of our calculations show that the regulatory outcome of CarD does in fact depend on the basal energy landscape (i.e. set of rate constants) of a promoter (Fig. 2*C*). More specifically, the model predicts that promoters with relatively unstable RP_o (i.e. rapid rates of bubble collapse and escape) will be activated by CarD whereas promoters with relatively stable RP_o (i.e. slow rates of bubble collapse and escape) will be repressed (Fig. 2*C*). Based on these analyses, we conclude that CarD has the potential to directly activate or repress transcript production as observed in the RNA-seq data.

Gene expression patterns in CarD mutants recapitulate changes in CarD's RPo stabilizing activity. Our flux-based modeling predicts that the transcriptional outcome of a CarD-regulated promoter depends on both the intrinsic RP_o stability and the RP_o stabilization afforded by CarD. Thus, for a given promoter, any mutation that impairs CarD's RP_o stabilizing activity should yield the same effect on transcript production. To test this prediction we compared the changes in expression profiles in each of the CarD mutants. Principal component analysis (PCA) of the *Mtb*

RNA-seq samples was performed and provided a general overview of how gene expression patterns among the CarD mutants clustered in relationship to each other. Sample replicates grouped tightly based on genotype, demonstrating inter-replicate consistency in our RNA-seq data (Fig. 3*A*). To further examine the gene expression relationships between samples, we performed sample-based hierarchical clustering based on Euclidean distance of VST (variance stabilizing transformation) transformed counts of all transcripts and found that the samples divided roughly into 3 groups (Fig. 3*B*). The first group included the CarD^{WT} replicates, the second group included the CarD^{R47E} and CarD^{K125A} replicates, and the third group included the replicates from the CarD^{I27F} and CarD^{I27W} strains.

Mutants from the same expression profile group showed differential expression of a similar subset of genes (hypergeometric test of over-enrichment $p<0.05$) (Fig. 3*C* and *D*). Furthermore, pairwise linear regression of expression fold-changes show that the direction and magnitude of gene expression changes within these groups showed a strong, positive correlation ($p<0.05$) (Fig. 3*C* and *D*). These data demonstrate that genes that are up- or down-regulated in CarD^{R47E} are likely to be similarly up- or down-regulated in CarD^{K125A}, and the same is true for CarD^{I27F} and CarD^{I27W}. These similarities in transcriptional profiles are consistent with data showing that CarD^{R47E} is phenotypically similar to CarD^{K125A} (97, 100) and CarD^{I27F} is phenotypically similar to CarD^{I27W} (101). In addition, the magnitudes of expression changes in CarD^{I27W} were larger than those of CarD^{I27F} (Fig. 3*D*), consistent with previously published data showing that the I27W mutation results in a larger increase in affinity to RNAP-□ than the I27F mutation (101). Thus, mutants with the same effect on CarD's RP_o-stabilizing activity group with each other based on gene expression patterns, suggesting that the gene expression changes observed in the mutant bacteria are a consequence of altered RP_o stability.

CarD mutants belonging to different expression profile groups also showed a significant overlap in the subsets of genes that were differentially expressed (hypergeometric test of over-enrichment $p<0.05$) (*SI Appendix* Fig. S2), despite not clustering together based on gene expression patterns (Fig. 3*B*). When we performed pairwise linear regressions, we did not observe a strong correlation, positive nor negative, when comparing the direction of gene expression changes of the genes significantly differentially expressed between the $CarD^{R47E}$/$CarD^{K125A}$ and the $CarD^{I27F}$/$CarD^{I27W}$ mutants (*SI Appendix* Fig. S2). These data demonstrate that mutations that impair (R47E/K125A) or enhance (I27F/I27W) CarD activity cause differential expression of a similar subset of *Mtb* genes, but the direction and magnitude of expression changes depend on how the specific mutation affects CarD's RP_o stabilizing activity.

Although pairwise comparisons revealed information about the gene expression profile relationships between the different CarD mutants, we sought to understand the patterns of expression of individual genes across the four CarD mutant strains. We chose to focus on genes that were most strongly affected by CarD mutation and examined a subset of 432 *Mtb* genes that were significantly differentially expressed 2-fold or greater in at least one mutant genotype. We performed gene-based hierarchical clustering of these 432 genes based on Pearson correlation of their $\log_2$ fold-change in expression compared to the $CarD^{WT}$ strain across the four CarD mutants and found that the genes separated into three clusters with two predominant expression trends (Fig. 3*E, SI Appendix* Dataset S2). Cluster 1 contained the largest number of genes (199/432), 127 of which were up-regulated in $CarD^{R47E}$ and $CarD^{K125A}$ and down-regulated in $CarD^{I27F}$ and $CarD^{I27W}$. Cluster 2 was the second largest (172/432) where most genes were down-regulated in $CarD^{R47E}$ and $CarD^{K125A}$. 65 of the 172 genes in cluster 2 were also up-regulated in $CarD^{I27F}$ and $CarD^{I27W}$. Almost all (198/199) of the genes in cluster 1 showed higher expression in $CarD^{R47E}$ and $CarD^{K125A}$

compared to CarD^{I27F} and CarD^{I27W}, whereas the opposite trend is true for almost all (171/172) of the genes in cluster 2. Therefore, clusters 1 and 2 reveal a prominent trend where mutants with impaired CarD activity (CarD^{R47E} and CarD^{K125A}) and mutants with enhanced CarD activity (CarD^{I27F} and CarD^{I27W}) show opposite changes in gene expression. Thus, many of the observed expression changes are likely the outcome of relatively stronger or weaker RP_o stabilizing activity by CarD.

The pairwise comparisons and the gene expression patterns in clusters 1 and 2 from the hierarchical clustering analysis demonstrate that the CarD^{R47E} and CarD^{K125A} mutants result in very similar transcriptional profiles. This indicates that despite facilitating distinct macromolecular interactions, CarD's RID and DBD usually work in concert to potentiate a single transcriptional outcome. Genes in cluster 3, which was the smallest of the three clusters (62/432), represent an exception to this relationship. Cluster 3 included a subset of 28 genes that were >2-fold up-regulated in CarD^{K125A} but either unchanged or down-regulated in CarD^{R47E}. This gene expression pattern is unusual given the positive correlation generally observed between CarD^{R47E} and CarD^{K125A} (Fig. 3C) and suggests that CarD's RID and DBD may have independent effects on transcription at a smaller subset of promoters. Nonetheless, our data indicate that in the majority of cases, the effect of a CarD mutant on gene expression outcome correlates with the effect of that mutation on CarD-mediated stabilization of RP_o, as predicted in the flux-based modeling.

Gene expression changes in CarD mutants correlate with DNA sequences in the -10 element and discriminator region of promoters. Our flux-based modeling predicts that the outcome of CarD activity on gene expression from different promoters depends on the intrinsic RP_o stability of the promoter. Since DNA sequences within a promoter recognition region (PRR) can influence the kinetics of transcription complex formation (129–131), we hypothesized that different

sequence elements within the PRR would correlate with the outcome of CarD activity. In *E. coli*, the -10 and -35 regions, the extended -10 element, UP elements upsteam of the -35 region, and the discriminator region between the -10 and TSS can affect RP_o stability (49, 113). Mycobacterial promoters contain highly variable sequences in the -35 region (60) and UP elements have not been reported. However, consensus "TANNNT" -10 sequence motifs and extended -10 sequences are prevalent in mycobacterial promoters (132) and have been shown to influence RP_o stability (133, 134).

To investigate whether promoter sequences correlated with the outcome of CarD activity, we used a published dataset of *Mtb* primary TSSs (132) and scanned a 50bp range upstream of the TSS for -10 element motifs. To avoid counting a promoter multiple times, we only considered genes that are directly downstream of a primary TSS. We assigned the *Mtb* promoters into one of four classes: extended consensus -10 "GNTANNNT", non-extended consensus -10 "TANNNT", extended non-consensus -10 "GNVANNNT", or "No Motif" if none of these motifs were detected within the -10 region. (*SI Appendix* Dataset S3). Across 1,778 primary TSS, 593 (33.3%) contained "GNTANNNT", 638 (35.9%) contained "TANNNT", 277 (15.6%) contained "GNVANNNT", and 270 (15.2%) contained "No Motif" (*SI Appendix* Fig. S3, Table S1), similar to what was previously reported (132). To determine whether there was a relationship between -10 promoter sequence and gene expression profiles in the CarD mutants, we compared the proportions of -10 classes between the differentially expressed gene clusters described in Figure 3E. We observed that clusters 1 and 3 were significantly enriched (hypergeometric test $p<0.05$) for promoters lacking a consensus -10 motif ("GNVANNT" and "No Motif") (Fig. 4*A*). Most of the genes in Cluster 1 were up-regulated in $CarD^{R47E}$ and $CarD^{K125A}$ and most of the genes in Cluster 3 were up-regulated in $CarD^{K125A}$ (Fig. 3*E*), suggesting that the direction of gene expression changes in these

mutants may also correlate with the -10 motif sequence. We examined the proportions of -10 classes for subsets of genes that were >2-fold significantly up- or down-regulated in CarD^{R47E} or CarD$^{K!25A}$ (*SI Appendix* Fig. S3*B*). Our promoter analysis revealed that genes up-regulated in CarD^{R47E} or CarD^{K125A} were significantly enriched for "Non-consensus" promoters lacking a "TANNNT" hexamer in their -10 region (Fig. 4*B*, hypergeometric test $p<0.05$). In contrast, the proportion of promoter motifs for genes down-regulated in CarD^{R47E} or CarD^{K125A} were similar to the genome as a whole. When we scanned the promoter regions for -35 elements, we were unable to identify consensus motifs, similar to other studies that have concluded that -35 elements in mycobacteria are highly variable (60, 135).

The GC content and length of the discriminator sequence between the TSS and the downstream edge of the -10 element has also been shown to influence transcription initiation kinetics (49). We examined the discriminator length and GC content for promoters significantly differentially expressed >2-fold in CarD^{R47E} or CarD^{K125A} (*SI Appendix* Table S2, Dataset S3). We omitted promoters lacking an identifiable -10 element ("No Motif"), because the discriminator could not be accurately positioned for these promoters. We found that genes up-regulated in CarD^{R47E} and CarD^{K125A} contained promoters with significantly shorter discriminator regions (Kruskal-Wallis $p<0.05$ followed by pairwise rank sum test) than promoters across the genome or promoters down-regulated in those mutants (Fig. 4*C*). In addition, the promoters of CarD^{R47E} and CarD^{K125A} up-regulated genes had a lower GC% in their discriminator regions (Fig. 4*D*).

Notably, a separate study by Shell *et al*. (136) mapped TSS in *Mtb* and identified numerous unique TSS that were not found by Cortes *et al*. (132). Although both datasets identified over 4,000 TSS, only 43.5% (2166/4798) TSS identified by Shell *et al*. were also mapped by Cortes *et al*. (*SI Appendix* Fig. S4). Therefore, we also performed our sequence analyses using the Shell *et al*.

dataset and observed similar trends (*SI Appendix* Fig. S5, Table S3. Collectively, our data suggests that genes that are up-regulated when CarD activity is impaired are enriched for promoters that lack a consensus -10 "TANNNT" hexamer and contain shorter, less GC-rich discriminator regions. These data support our model that DNA sequences in the promoter that influence intrinsic RP_o stability may also influence the transcriptional outcome of CarD activity.

Discussion

Previous mechanistic studies of CarD's function on mycobacterial promoters have been mostly limited to the *Mtb* r*rnA*P3 promoter. However, the *rrnA*P3 promoter is similar in architecture to the consensus *E. coli* promoter and is not representative of most *Mtb* promoters, which are notably diverse in structure (60). In this study, we generated RNA-seq data using four different CarD mutant *Mtb* strains to investigate CarD's activity in the context of native *Mtb* promoter sequences genome-wide. Our results show that when CarD activity is perturbed, approximately two thirds of the *Mtb* genome is differentially expressed, highlighting CarD's broad importance for coordinating *Mtb* gene expression. The genes affected by mutation of CarD fell into multiple diverse functional categories, indicating that CarD does not regulate a specific regulon of genes like a traditional transcription factor but rather participates in transcription initiation as a basic part of the mycobacterial RNAP machinery. Therefore, CarD's essentiality in *Mtb* is likely due to its broad importance in maintaining global gene expression patterns that support viability.

We also discovered that mutating CarD results in a roughly equal number of up-regulated and down-regulated genes, indicating that CarD is capable of activating and repressing transcription from different *Mtb* promoters. The ability of CarD to repress gene expression had not

been appreciated in previous studies with the *rrnA*P3 promoter. In principle, this observation could be explained by indirect effects where CarD enhances the expression of a transcription factor that represses the expression of a set of genes. However, there is also precedent for factors that directly modify RP_o stability being able to up-regulate the expression of one gene and down-regulate another depending on the kinetic properties of the transcription initiation complexes at that promoter (92, 95, 127). Our previous kinetic model of CarD activity only considered CarD's effect on the rates of RP_o formation and RP_o collapse (46) and could not adequately explain how CarD's RP_o stabilizing activity could result in differential transcription activation or repression. Here, we show that CarD slows promoter escape from the *rrnAP3* promoter, which we hypothesize occurs via the additional RNAP-promoter DNA interactions that result from CarD association. By including CarD's ability to inhibit promoter escape in a flux-based model of CarD's effect on transcript production (127), our measurements support that stabilization of RP_o by CarD could lead to up and down-regulation of gene expression, thus explaining the RNA-seq data. Our data comparing mutations that impair CarD's ability to stabilize RP_o (R47E and K125A) with mutations that enhance CarD's activity (I27F and I27W) (46, 101) supported this model where mutations with opposite effects on RP_o stability often resulted in opposite changes in gene expression.

Our flux-based model implies that the outcome of CarD's activity at a promoter depends on intrinsic kinetics of the transcription initiation complexes that form at a particular promoter. In *E. coli*, RP_o stability and promoter escape rates are highly dependent upon sequences in the promoter recognition region (PRR) (49, 129, 137). In addition, the small alarmone ppGpp binds to RNAP in *E. coli* to elicit bi-directional transcriptional responses by altering initiation kinetics (89), and the outcome of ppGpp activity is correlated with certain promoter DNA sequences (95). Most promoters within the GC-rich mycobacterial species do not resemble *E. coli* promoters (60,

133, 135, 138) and the impact of PRR sequences in *Mtb* on transcription initiation kinetics has yet to be investigated in detail. We found that genes up-regulated in CarD^{R47E} and CarD^{K125A} were significantly enriched for promoters lacking a consensus "TANNNT" motif in their -10 element and contain shorter, less GC-rich discriminator regions. Based on our data, none of these sequence elements are the sole determinant of the outcome of CarD activity because many promoters containing consensus -10 elements and longer GC-rich discriminator regions were also up-regulated, and many non-consensus promoters were also down-regulated. Therefore, a combination of sequence elements likely affects the outcome of CarD activity, dependent on the overall impact the PRR sequences on transcription initiation kinetics. A caveat of our analyses based on *in vivo* expression profiles is that genes that are indirectly differentially expressed as a result of CarD mutation may be masking our ability to discern specific promoter sequences that contribute to the outcome of CarD activity. Future work will focus on performing *in vitro* assays to isolate only the direct effects of CarD activity on transcription from different promoter sequences.

A previous study suggested that a thymine base at the -12 position (T_{-12}) within the -10 element is important for CarD's conserved W85 residue to inhibit bubble collapse (96). The T_{-12} identified in that study likely corresponds to the upstream thymine residue found in consensus "TANNNT" -10 elements. Our promoter analysis suggests that "Non-consensus" promoters lacking this thymine are enriched in the genes up-regulated in *Mtb* mutants with impaired CarD activity (Fig. 4*A-B*), suggesting that a T_{-12} is not required for CarD to affect expression. However, given that the *Mtb* CarD^{W85A} mutant is not viable (100), we were unable to directly interrogate the importance of this interaction between W85 and the upstream thymine in the conserved -10 hexamer.

In addition to cis-acting PRR sequences, trans-acting factors, such as other transcription factors, could also contribute to the effect of CarD on transcription of a particular gene. *Mtb* encodes another essential RNAP-interacting protein, RbpA (70, 73), which is also capable of stabilizing RP_o formed by mycobacterial RNAP *in vitro* (28, 64, 74). Structural studies have shown that association of CarD and RbpA with the same RNAP holoenzyme is feasible (26, 28) and we have shown that CarD and RbpA can function cooperatively to stabilize RP_o (64). In addition to CarD and RbpA, *Mtb* also encodes numerous two-component systems and other transcription factors that allow the bacteria to alter its gene expression in response to environmental cues (68). Future studies will focus on elucidating how RbpA and other transcriptional regulators impact on the outcome of CarD activity at particular promoters.

Collectively, our data show that CarD coordinates global gene expression patterns in *Mtb* by activating and repressing transcription via stabilization of RNAP-promoter complexes. A recent study highlighted the potential of prokaryotic transcription factors with single regulatory mechanisms, such as CarD, to potentiate diverse regulatory outcomes in a promoter-specific manner (127). In this manuscript we demonstrate *in vivo* that CarD activity leads to diverse gene expression outcomes and does not solely act as a transcriptional activator. Importantly, CarD homologs are conserved in many bacterial phyla (63, 68) and regulate transcription to mediate diverse processes including stress responses, cell division, and metabolic homeostasis (104, 107, 109, 116). Therefore, these studies shed light on the maintenance of optimal gene expression patterns broadly across bacteria.

Acknowledgements

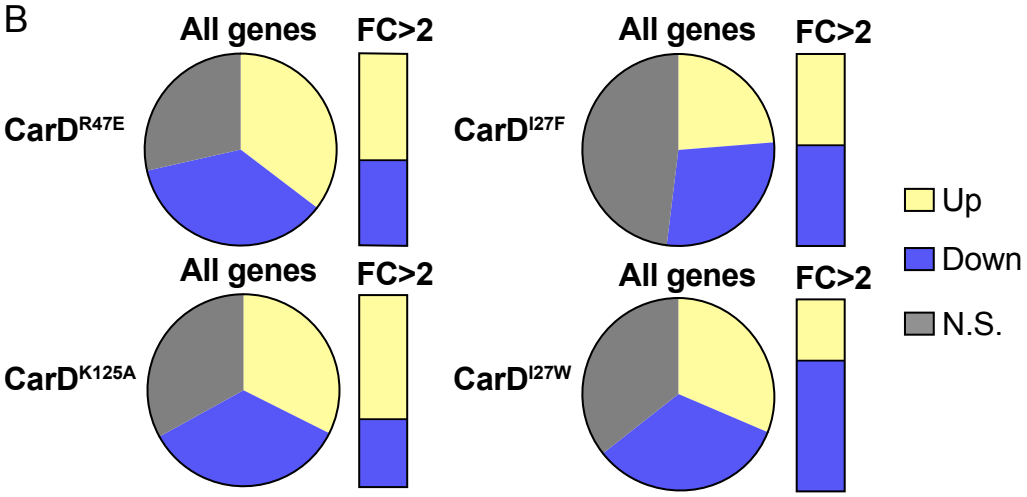
This work was supported by Grant GM107544 from the National Institutes of Health (NIH) (C.L.S. and E.A.G.), NIH NIAID Award T32A1007172 (D.Z.), NIGMS Cell and Molecular Biology Training Grant GM007067 (A.L.G.), and the Stephen I. Morse Graduate Fellowship (A.L.G.). C.L.S. is also supported by a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease Award. Sequencing was performed at the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine. 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 (NCRR), a component of the NIH, and NIH Roadmap for Medical Research.

Figures

A

Mutation	Domain	Effect
R47E	RID	Decreases interaction strength with RNAP
K125A	DBD	Mutation in the DNA-binding basic patch
I27F	RID	Increases interaction strength with RNAP
I27W	RID	Increases interaction strength with RNAP

B



		CarD^{R47E}	CarD^{K125A}	CarD^{I27F}	CarD^{I27W}
All genes	**Up**	35% (1420)	32% (1301)	24% (953)	31% (1243)
	Down	36% (1457)	35% (1389)	28%(1133)	33% (1343)
FC>2	**Up**	3.9% (155)	2.8% (113)	0.90% (36)	1.1% (45)
	Down	3.1% (123)	1.5% (62)	0.92% (37)	2.1% (83)

Figure 1. *Mtb* strains encoding point mutants of CarD show genome-wide differences in gene expression. (*A*) Diagram showing the protein domain structure of CarD, which consists of a N-terminal RNAP-interacting domain (RID; amino acids 1-66) and a C-terminal DNA binding domain (DBD; amino acids 67-162). The table lists the four CarD point mutations that were used in this study and their effects on CarD function. (*B*) Pie charts display the proportion of 4016 *Mtb*

protein encoding genes with nonzero reads that were significantly (p_{adj}<0.05) up-regulated (fold-change>0), down-regulated (fold-change<0), or not significantly differentially expressed (N.S., p_{adj}≥0.05) in each CarD mutant relative to their expression in $CarD^{WT}$. The bars under FC>2 shows the distribution of up- and down-regulated genes for the subset of *Mtb* genes that were significantly differentially expressed greater than two-fold in each mutant genotype. These data are enumerated in the table below. Differential expression testing was performed using RNA-seq data from three replicates of each genotype.

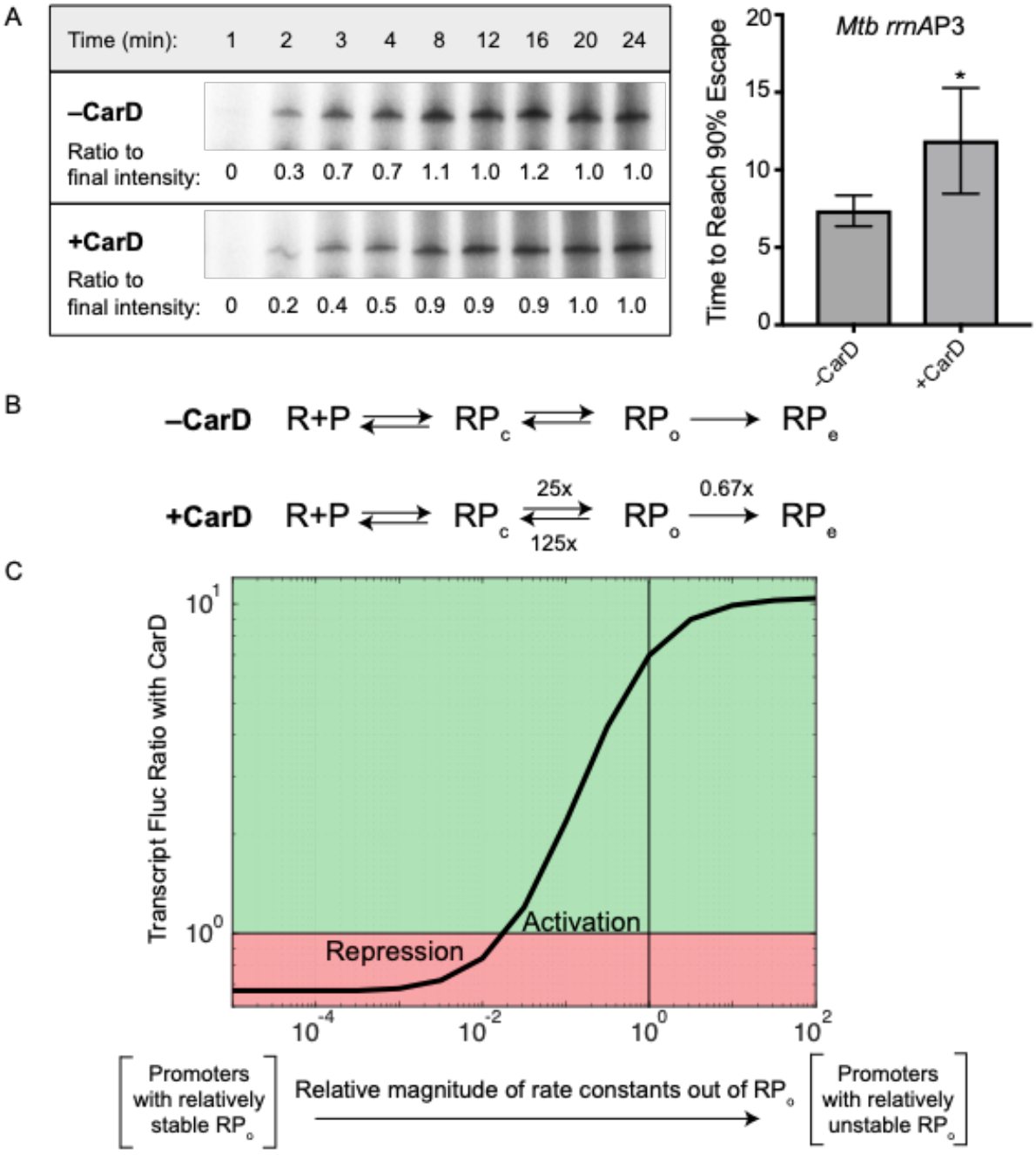


Figure 2. CarD slows promoter escape from the *Mtb rrn*AP3 promoter. (*A*) Single-round *in vitro* promoter escape assay results with representative gel images showing the time-dependent increase in ^{32}P-labeled RNA transcripts formed by *M. bovis* RNAP from the *Mtb rrn*AP3 promoter construct in the presence and absence of CarD. Promoter escape rate is quantified by the time until 90% of

the final transcript intensity is reached ($t_{90\%}$). The graph shows the mean $t_{90\%}$ ± SEM (-CarD n=3, +CarD n=7). Statistical significance was analyzed by Welch's t-test. *, $p<0.05$. (*B*) A proposed kinetic model in which CarD accelerates the rate of transcription bubble formation ($RP_c \rightarrow RP_o$) 25-fold, slows the rate of bubble collapse ($RP_o \rightarrow RP_c$) 8-fold, and slows the rate of promoter escape ($RP_o \rightarrow RP_e$) 1.5-fold. Effects of CarD on rates were chosen based on experimentally determined values. (*C*) Graph showing the mRNA flux ratio from a given *Mtb* promoter upon the addition of CarD. The X-axis represents a titration of the rate constants out of RP_o relative to the kinetic model for *rrnA*P3. Calculations were performed with the kinetic model of CarD activity on a set of hypothetical promoters with a titration of RP_o stability using the web-based tool described in (127). The green region represents promoters that would be activated by CarD while the red region represents promoters that would be repressed.

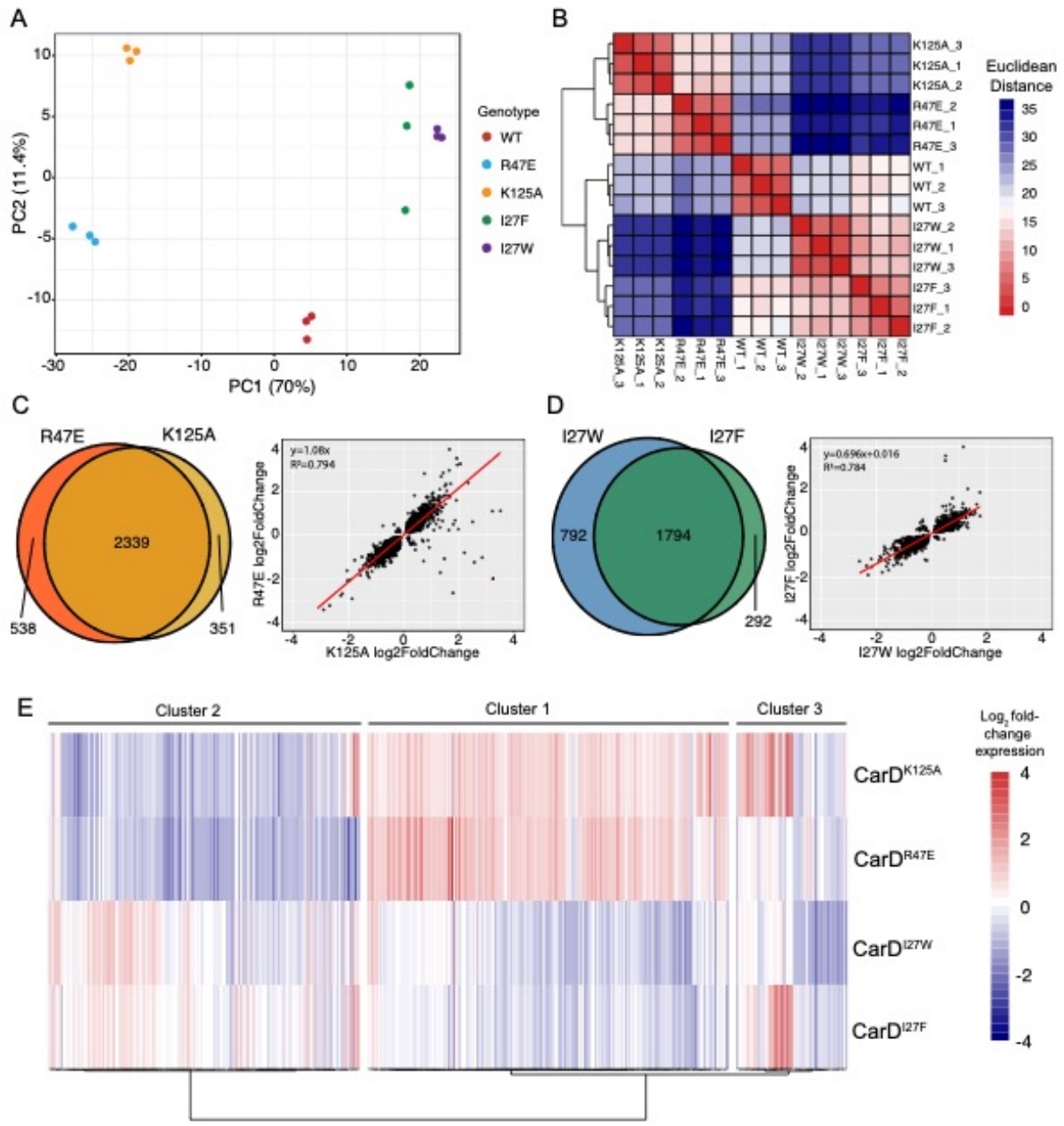


Figure 3. CarD mutants with similar effects on RP_o stability show similar gene expression profiles. (*A*) A principal component analysis (PCA) of RNA sequencing (RNAseq) samples based on read counts generated by mapping reads to a library of protein encoding sequences from the *Mtb H37Rv* genome. Each point represents one sequencing sample colored by genotype, and the distance

between two points reflects the variance in gene expression between the two samples. The first two principle components PC1 and PC2 define the x- and y-axes, respectively, and account for 70% and 11.4% of the variance, respectively. (*B*) Hierarchical clustering of RNAseq samples based on read count data where each column and row represents one sample. The color of each cell represents the Euclidean distance, calculated based on relative expression of all *Mtb* protein encoding genes, between each sample pair. (*C-D*) Venn diagrams and scatter plots show the overlap in the lists of *Mtb* genes that were significantly differentially expressed ($p_{adj}<0.05$) between (*C*) $CarD^{R47E}$ and $CarD^{K125A}$, and (*D*) $CarD^{I27F}$ and $CarD^{I27W}$. Scatter plots compare gene expression changes for each pair of mutants. Each point represents an *Mtb* gene that was significantly differentially expressed in one or both of the mutant strains being compared. The position of the dot along the axes represents the $\log_2$ fold change relative to $CarD^{WT}$, with one mutant on the x-axis and the other on the y-axis. Red lines represent a linear regression line calculated for each pair. (*E*) Unsupervised hierarchical clustering (Pearson distance, Ward's method linkage) of patterns of gene expression changes in a subset of 432 *Mtb* genes that were significantly differentially expressed greater than two-fold in at least one CarD mutant *Mtb* strain relative to $CarD^{WT}$. Each column represents a different gene in the set and each row represents a different CarD mutant strain. The coloring of each cell shows the $\log_2$ fold change in expression for each gene in a given CarD mutant strain relative to $CarD^{WT}$

.

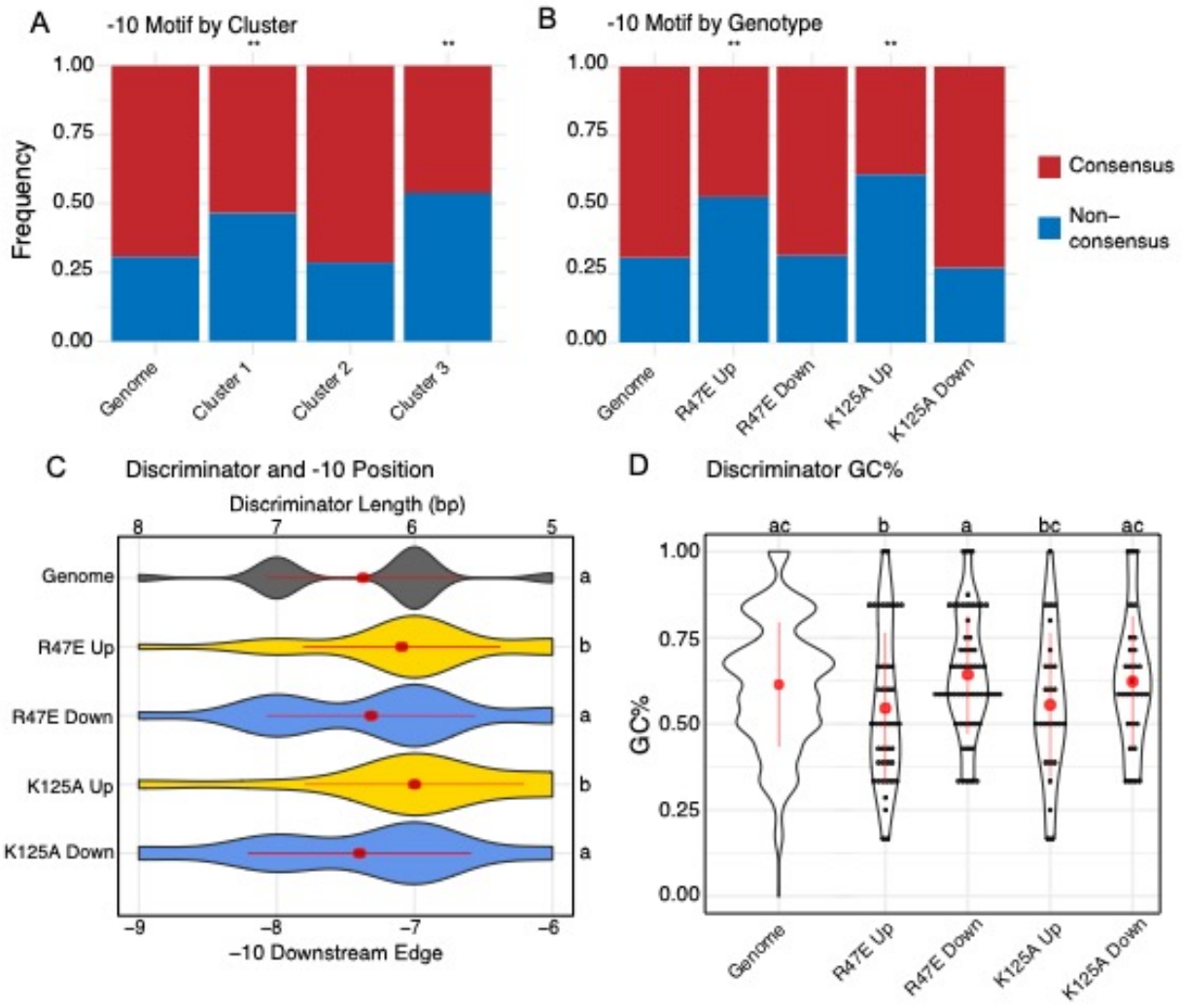


Figure 4. Genes that are up-regulated in RP_o-destabilizing CarD mutant *Mtb* strains are associated with promoters lacking a consensus "TANNNT" -10 element motif and containing shorter and less GC-rich discriminator. *Mtb* genes directly downstream of a primary TSS identified in Cortes *et al*. (132) were classified based on the presence of "TANNNT" DNA sequence motif in their promoter -10 element. (A) and (B) The relative proportions of these promoter classes are shown for the entire *Mtb* genome ("Genome"), (*A*) in gene clusters defined in figure 3*E*, and (*B*) in subsets of genes that were significantly up- or down-regulated in $CarD^{R47E}$ or $CarD^{K125A}$. "Non-consensus" promoters were significantly overrepresented (hypergeometric test; *=p<0.05, **=p<0.005,

n.s.=not significant) in genes up-regulated in CarD^{R47E} and CarD^{K125A}. (*C*) Violin plots showing the distribution of discriminator lengths for *Mtb* promoter subsets. Promoters without an identifiable -10 element were not included. (*D*) Violin and dot plots showing distribution of discriminator GC% for *Mtb* promoter subsets. Red dots and lines represent mean discriminator length or GC% ± SD. Statistically significant differences in discriminator length and GC% were detected using a Kruskal-Wallis rank sum test followed by post-hoc pairwise Dunn's Tests where groups with different letters (a, b, or c) are significantly different from each other (unadjusted $p<0.05$). Genome n=1778, R47E Up n=55, R47E Down n=73, K125A Up n=46, K125A Down n=37; Cluster 1 n=88, Cluster 2 n= 91, Cluster 3 n=26.

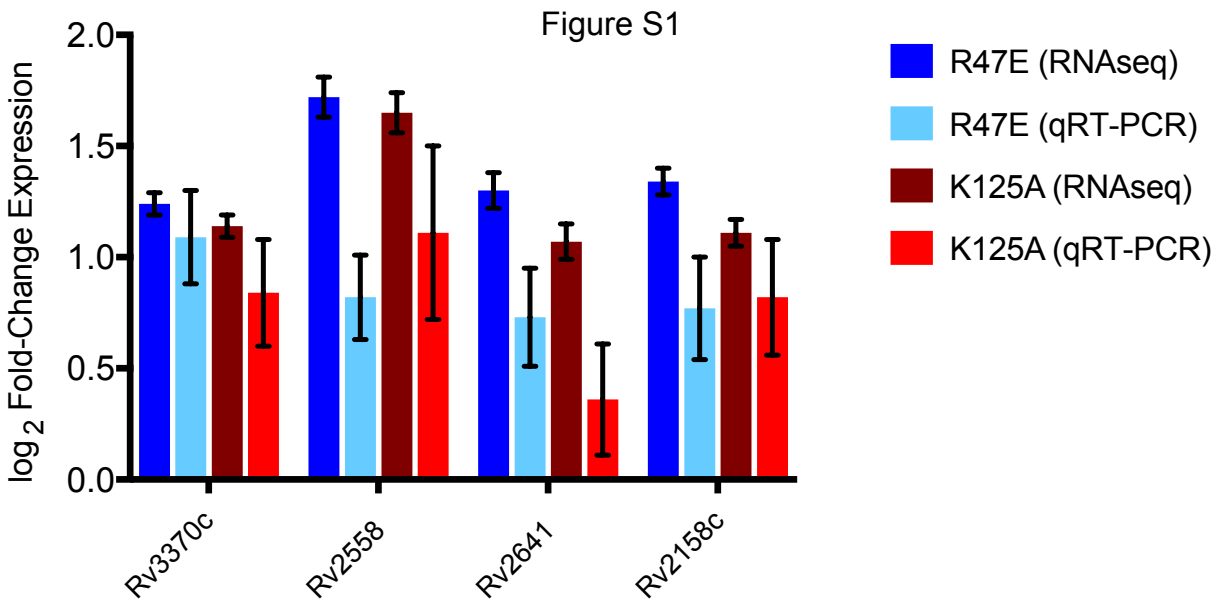


Figure. S1 Bar graphs comparing the fold-change in expression in CarD^{47E} and CarD^{K125A} relative to CarD^{WT} as measured by quantitative reverse transcription PCR (qRT-PCR) with RNA spike in controls (n=5) or RNA-seq (n=3) for four *Mtb* genes that were highly up-regulated according to RNA-seq. For qRT-PCR values, the transcript abundance for each *Mtb* gene is normalized to the abundance of MS2 phage RNA (Roche 10165948001), which is added to each sample in proportion to cell number. The graph shows mean ± SEM.

Chapter 3: Transcription regulation by CarD in mycobacteria is guided by basal promoter kinetics

Dennis X. Zhu, Christina L. Stallings

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

Abstract

Bacterial pathogens like *Mycobacterium tuberculosis* (*Mtb*) employ transcription factors to adapt their physiology to the diverse environments within their host. CarD is a conserved bacterial transcription factor that is essential for viability in *Mtb*. Unlike classical transcription factors that recognize promoters by binding to specific DNA sequence motifs, CarD binds directly to the RNA polymerase (RNAP) to stabilize the open complex intermediate (RP_o) during transcription initiation. We previously showed using RNA-sequencing that CarD is capable of both activating and repressing transcription *in vivo*. However, it is unknown how CarD achieves promoter specific regulatory outcomes in *Mtb* despite binding indiscriminate of DNA sequence. We propose a model where CarD's regulatory outcome depends on the promoter's basal RP_o stability and test this model using *in vitro* transcription from a panel of promoters with varying levels of RP_o stability. We show that CarD directly activates full-length transcript production from the *Mtb* ribosomal RNA promoter *rrnA*P3 (AP3) and that the degree of transcription activation by CarD is negatively correlated with RP_o stability. Using targeted mutations in the extended -10 and discriminator region of AP3, we show that CarD directly represses transcription from promoters that form relatively stable RP_o. DNA supercoiling also influenced RP_o stability and affected the direction of CarD regulation, indicating that the outcome of CarD activity can be regulated by factors beyond promoter sequence. Our results provide experimental evidence for how RNAP-binding transcription factors like CarD can exert specific regulatory outcomes based on the kinetic properties of a promoter.

Introduction

Throughout their life cycle, bacteria must continuously adapt their physiology to respond to and survive in their changing environments. As such, the ability to sense environmental signals and transduce these cues into an appropriate physiological response is important for the virulence of pathogens such as *Mycobacterium tuberculosis* (*Mtb*), which face threats from both the host immune system and antibiotic treatment. Regulation of transcription initiation is a major mechanism by which bacteria adapt their gene expression in response to environmental stimuli. Transcription in bacteria is performed by a single RNA polymerase (RNAP) enzyme, which consists of a multi-subunit core enzyme that can bind to different sigma factors (σ) to form a holoenzyme and initiate promoter-specific transcription. *Mtb* devotes a significant fraction of its genome towards encoding numerous transcription factors that can regulate transcription initiation by altering the promoter specificity and recruitment of RNAP (68, 80). Classically, transcription factors are recruited to promoters by recognizing and binding a DNA sequence motif, which allows the factor to specifically regulate a subset of the genome. However, some bacteria also encode transcription factors that instead localize to promoter regions by binding directly to RNAP (69, 86). This class of transcription factors is best exemplified by the stringent response regulators DksA and guanosine (penta)tetraphosphate [(p)ppGpp], which bind to the *Escherichia coli* RNAP to directly activate or repress transcription from subsets of *E. coli* promoters (139). These factors exert promoter specific transcription regulation despite being unable to discriminate promoters at the level of binding. The prevailing hypothesis for the mechanism behind this promoter specificity postulates that these factors can potentiate different outcomes on transcription depending on the underlying initiation kinetics of a promoter (92). Recently, this hypothesis has also been applied

to the regulatory mechanisms of other RNAP-binding transcription factors such as CarD (111, 127).

CarD is an RNAP-binding transcription regulator that is widely conserved across many eubacteria phyla and essential for viability in mycobacteria (63). CarD associates with transcription initiation complexes by binding directly to the RNAP β subunit through its N-terminal RNAP-interaction domain (RID) (63, 97). The CarD C-terminal DNA-binding domain (DBD) also interacts with DNA at the upstream fork of the transcription bubble in a sequence-independent manner (26, 96, 100). Numerous kinetic studies have demonstrated that CarD stabilizes the RNAP-promoter open complex (RP_o) formed by the mycobacterial RNAP during transcription initiation (25, 28, 46, 59, 100). CarD accomplishes this through a two-tiered kinetic mechanism in which it binds to RNAP-promoter closed complexes (RP_c) to increase the rate of DNA melting while also slowing the rate of bubble collapse (46). Furthermore, by stabilizing RP_o, CarD slows the rate of promoter escape (75), which is a necessary step preceding full-length RNA synthesis. Due to its ability to stabilize RP_o *in vitro*, it was expected that CarD functioned generally as a transcription activator. However, although numerous studies have examined CarD's effect on individual rate constants between transcription initiation intermediates (28, 46, 75), the composite effect of CarD's kinetic mechanism on full-length RNA production remains unknown. Furthermore, while *in vitro* studies of CarD have utilized only a handful of promoters, primarily focusing on the *Mtb* ribosomal RNA promoter *rrnA*P3 (AP3) (28, 46, 59, 64, 75, 96, 100, 101), chromatin immunoprecipitation sequencing (ChIP-seq) in *Mycobacterium smegmatis* indicates that CarD co-localizes with the housekeeping sigma factor σ^A to promoter regions broadly across the mycobacterial genome (25, 103), leaving a gap in our understanding of CarD's activity under different promoter contexts.

To characterize CarD's role in transcription regulation throughout the mycobacterial genome, we previously performed RNA-sequencing (RNA-seq) on a set of *Mtb* strains expressing mutants of CarD that either impair or enhance its ability to stabilize RP_o *in vitro* (111). We discovered that altering CarD activity in *Mtb* led to both up-regulation and down-regulation of numerous protein-encoding transcripts, suggesting that CarD could function as either a transcriptional activator or a transcriptional repressor in different promoter contexts. Prior *in vitro* studies with *Rhodobacter sphaeroides* CarD and RNAP have shown that *Rsp*CarD activates transcription from promoters lacking a conserved T at the -7 position (110) and represses transcription from its own promoter (140). However, unlike Alphaproteobacteria like *R. sphaeroides*, which contain a T_{-7} at fewer than 50% of their promoters, most other bacterial phyla, including Actinobacteria like *Mtb*, have a T_{-7} at over 90% of their promoters (110), making it unlikely that the T_{-7} is a conserved mechanism of CarD promoter specificity. Instead, we previously proposed a model in which the outcome of CarD regulation is dependent on the basal transcription initiation kinetics at a given promoter (111, 127). Specifically, at unstable promoters that are rate-limited at the step of bubble opening, CarD would facilitate full-length RNA production by stabilizing RP_o, while at stable promoters that are rate-limited at the step of promoter escape, CarD would make it more difficult for RNAP core enzyme to break contacts with promoter DNA. Herein, we directly test our model using *in vitro* transcription approaches to explore the relationship between RP_o stability and transcription regulation by CarD. We discover that both promoter DNA sequence and DNA topology influence the basal RP_o stability of a promoter and the regulatory outcome of CarD on transcription. In addition, we find that in the context of a promoter with high basal RP_o stability, CarD can directly repress transcription, marking the first demonstration of direct transcriptional repression by *Mtb* CarD. This work provides experimental

evidence for how RNAP-binding transcription factors like CarD can potentiate multiple regulatory outcomes on transcription through a single kinetic mechanism.

Experimental Procedures

Bacterial growth and RNA collection

All *M. smegmatis* strains used in this study were derived from mc²155 and grown in LB medium supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% tween-80 at 37 °C. *M. smegmatis* strains expressing $CarD^{WT}$, $CarD^{R25E}$, $CarD^{K125E}$, or $CarD^{I27W}$ were engineered so that the native copy of *carD* is deleted, and the respective CarD allele is expressed from a constitutive P*myc1-tetO* promoter integrated into the genome. The construction of these strains has been previously described (100, 101). For RNA collection, *M. smegmatis* cultures were grown to OD_{600} 0.5-0.9, pelleted, and lysed in TRIzol reagent (Invitrogen) by bead-beating. RNA was isolated by TRIzol-chloroform extraction followed by isopropanol precipitation and finally resuspended in nuclease-free water (Invitrogen).

RNA sequencing and data analysis

RNA samples were DNase treated using the TURBO DNA-*free* Kit (Invitrogen) and submitted to the Washington University Genome Technology Access Center for paired-end Illumina sequencing (NovaSeq 6000 XP). Ribosomal RNA was depleted prior to sequencing using the Qiagen FastSelect system. Illumina reads were pre-processed using *FastQC* and adapter sequences were removed using *trimmomatic* (141). Sequencing reads were aligned using *HiSat2* (142) to the *M. smegmatis* mc²155 reference genome (assembly ASM1500v1) from the *Ensembl* database (143). Reads mapping to annotated protein coding regions were quantified using *featureCounts*

(144). Differential expression analysis was performed using *DESeq2* (118). Downstream data analysis and visualization was performed using custom R scripts.

Protein purification

Plasmids containing the *M. tuberculosis* H37Rv genomic DNA encoding the different *Mtb* RNAP holoenzyme subunits were a gift from Jayanta Mukhopadhyay (Bose Institute, Kolkata, India). *Mtb*RNAP-σ^A holoenzyme was purified as previously described (74, 145). Briefly, *Mtb* *Mtb*RNAP-σ^A holoenzyme protein was expressed in *E. coli* BL21 cells containing the plasmids pET-Duet-*rpoB-rpoC* (encoding the β and β' subunits), pAcYc-Duet-*sigA-rpoA* (encoding an N-terminal 10xHis-tagged-σ^A subunit and α subunits), and pCDF-*rpoZ* (encoding the ω subunit). Holoenzyme protein was isolated from *E. coli* cell lysate by affinity chromatography using a 2x 5mL HisTrap HP Ni^{2+} affinity columns (Cytiva) and further purified by size exclusion chromatography using a Sephacryl S-300 HiPrep column (Cytiva) to select for associated holoenzyme. Purified *Mtb*RNAP-σ^A holoenzyme was flash frozen in storage buffer (50% glycerol, 10mM Tris pH 7.9, 200mM NaCl, 0.1 mM EDTA, 1mM $MgCl_2$, 20μM ZnCl, and 2mM DTT) and stored at -80 °C. CarD proteins were expressed in BL21 *E. coli* cells using the pET SUMO vector system described previously (74). Purified CarD protein was stored in 20mM Tris pH 7.9, 150mM NaCl, and 1mM beta-mercaptoethanol.

***In vitro* transcription**

Promoter fragments used for *in vitro* transcription were prepared by annealing two complementary single-stranded DNA oligos (IDT) containing the WT or variant AP3 promoter sequence from positions -39 to +4 relative to the transcription start site to create a linear double-stranded DNA fragment that was ligated into the pMSG434 plasmid. Linear DNA templates used for *in vitro*

transcription were prepared PCR amplifying a 437bp fragment from the pMSG434 plasmid. Plasmid DNA templates for *in vitro* transcription were constructed by inserting an intrinsic transcription termination sequence (5'-TTTAT-3') into the pMSG434 plasmid 70bp downstream of the cloned AP3 transcription start site. Negatively supercoiled plasmids were grown in *E. coli* and then isolated using a QIAGEN Plasmid Midi Kit. To generate cut or nicked plasmid templates, plasmid DNA was incubated with XhoI restriction endonuclease (NEB) at 37 °C or Bt.NstNBI nicking endonuclease (NEB) at 55 °C for 1 hour, respectively. All DNA templates were purified by extracting with Buffer-Saturated Phenol pH >7.4 (Invitrogen) followed by isopropanol precipitation before being used in *in vitro* transcription reactions. A full list of the primers used to construct the DNA templates can be found in **Table S6**.

Multi-round *in vitro* transcription assays were performed by combining *Mtb*RNAP-σ^A holoenzyme, template DNA, and NTPs in a 20μL reaction volume. Multi-round reactions contained final concentrations of 40nM RNAP holoenzyme, 0.8nM DNA template, 0.1mg/mL BSA, 1mM DTT, 400μM GTP, 200μM ATP, 200μM CTP, 200μM UTP, 20μCi/mL [α-^{32}P]-UTP (PerkinElmer), 10mM Tris-HCl pH 7.9, 10mM $MgCl_2$, and 40mM NaCl. Reactions were initiated with the addition of NTPs and incubated at 37 °C for 1 hour before being terminated with the addition of 20μL 'stop buffer' (95% formamide, and <0.1% bromophenol blue and xylene cyanol). Three nucleotide *in vitro* transcription reactions were performed in the same manner, except with final reaction concentrations of 100nM RNAP holoenzyme, 10nM DNA template, 0.1mg/mL BSA, 1mM DTT, 20μM GpU, 10μM UTP, 62.5μCi/mL [α-^{32}P]-UTP, 10mM Tris-HCl pH 7.9, 10mM $MgCl_2$, and 40mM NaCl. Multi-round and three nucleotide *in vitro* transcription reaction products were separated by gel electrophoresis on denaturing (7M urea) 8% or 22% polyacrylamide gels, respectively, which were vacuum dried and visualized using a

phosphorimager screen. Reactions with CarD contained 25:1 molar ratio CarD:RNAP (1μM CarD for the multi-round *in vitro* transcription reactions or 2.5 μM CarD for the three nucleotide transcription reactions) unless otherwise noted.

Results

CarD binding correlates with transcriptional regulation but not the direction of regulatory outcome

A fundamental feature in our model of CarD mechanism is that the regulatory outcome of CarD on a given mycobacterial promoter is determined based on differences in the basal transcription initiation kinetics of the promoter and not differences in CarD binding. This model is based on comparing ChIP-seq data from *M. smegmatis*, where CarD is present at almost all RNAP-σ^A transcription initiation complexes (25, 103), with RNA-seq data from *Mtb*, where mutation of CarD resulted in both up- and downregulation of gene expression (111). However, we cannot yet rule out the alternative hypothesis that CarD's uniform localization pattern in *M. smegmatis* represents a unidirectional transcription activating mechanism for CarD in *M. smegmatis* in contrast to the bi-directional regulatory activity in *Mtb* that is suggested by our RNA-seq data.

To address this gap in our model, we performed an RNA-seq experiment in *M. smegmatis* that could be directly compared to the *M. smegmatis* ChIP-seq dataset. In our published *Mtb* RNA-seq experiment (111), we collected RNA from *Mtb* strains expressing mutant alleles of CarD with either weakened affinity for RNAP ($CarD^{R47E}$), predicted weakened affinity for DNA ($CarD^{K125A}$), or increased affinity for RNAP ($CarD^{I27F}$ and $CarD^{I27W}$). By collecting RNA from *Mtb* strains with mutations that target different domains of CarD, we were able to dissect how the respective

interactions of CarD's functional domains contributed to its role in regulating the *Mtb* transcriptome. To replicate this experimental design in *M. smegmatis*, we collected RNA from four strains of *M. smegmatis* with the native copy of *carD* deleted and expressing one of four different alleles of *Mtb* CarD: wild-type (WT) CarD ($CarD^{WT}$), $CarD^{R25E}$ (a RID mutant with weakened affinity for RNAP), $CarD^{K125E}$ (a DBD mutant with weakened affinity for DNA), or $CarD^{I27W}$ (a RID mutant with increased affinity to RNAP) as the only *carD* allele. Similar to the CarD mutations used in our *Mtb* experiment, the CarD mutations that weaken its macromolecular interactions with RNAP or DNA (R25E and K125E) impair CarD's ability to stabilize RP_o *in vitro* (46, 100), while the I27W mutation increases its affinity for RNAP and allows CarD to potentiate RP_o-stabilization at lower concentrations (101). For each strain, we collected RNA from four biological replicates of exponentially growing cells in nutrient replete conditions for sequencing. Two replicates ($CarD^{R25E}$-1 and $CarD^{K125E}$-4) were identified as outliers following principal component analysis (PCA) and were discarded from downstream analysis (**Fig. S1**).

In all three strains with mutations in *carD,* over 25% of the 6,716 coding genes in *M. smegmatis* Mc^2155 were significantly differentially expressed ($P_{adj} < 0.05$) in comparison to the $CarD^{WT}$ strain (**Fig. 1A**). The number of differentially expressed genes in the $CarD^{R25E}$ (2909 genes) and $CarD^{K125E}$ (2901 genes) *M. smegmatis* strains is similar to the number of differentially expressed genes in the $CarD^{R47E}$ (2877 genes) and $CarD^{K125A}$ (2690 genes) *Mtb* strains (111). However, homologous genes between the two species showed little correlation in their transcript expression patterns (**Fig. S2**), suggesting that CarD does not simply regulate a subset of homologous genes conserved between *Mtb* and *M. smegmatis*. Each of the *M. smegmatis* CarD mutant strains exhibited a similar number of up-regulated genes as down-regulated genes (**Fig. 1A**), following the same pattern as the *Mtb* CarD mutant strains (111) and suggesting that CarD is

capable of potentiating both transcriptional activation and repression in *M. smegmatis*. Importantly, the strains did not show significant differences in the total amount of RNA per cell (**Fig. 1B**), suggesting that the transcript abundance differences measured in the CarD mutant strains represent local changes in transcription at specific genes rather than a global decrease in RNA production within the cell that would be expected if CarD functioned strictly as a transcriptional activator.

The transcriptomic relationship between different CarD mutant strains in *M. smegmatis* was also consistent with the relationships we observed in our *Mtb* dataset (111). PCA of the RNAseq data illustrated that the *M. smegmatis* sample replicates clustered tightly with each other based on CarD genotype and samples from CarD mutant strains with impaired RP_o-stabilizing activity ($CarD^{R25E}$ and $CarD^{K125E}$) separated from the strain with enhanced RP_o-stabilizing activity ($CarD^{I27W}$) along the first principal component (**Fig. 1C**), demonstrating consistency between replicates from the same genotype and suggesting that altered CarD RP_o-stabilizing activity contributes to transcript abundance changes in the mutant bacteria. In addition, the $CarD^{R25E}$ and $CarD^{I27W}$ strains, which encode CarD RID mutants with impaired or enhanced RP_o-stabilization *in vitro*, respectively, displayed largely opposite transcriptomic changes (**Fig. S3A**), similar to the RID mutants in *Mtb* (111). In the PCA, the $CarD^{K125E}$ samples separated from all other samples along the second principal component (**Fig. 1A**) and the direction of transcript abundance changes in the DBD mutant $CarD^{K125E}$ samples correlated poorly with the transcript abundance changes in the RID mutant $CarD^{R25E}$ samples ($R^2 = 0.351$) (**Fig. S3B-C**). This is in contrast to the tight correlation between CarD RID and DBD mutants in *Mtb* (111) and may suggest that mutations in the DBD and RID have unique effects on CarD's regulatory function in *M. smegmatis*.

The ChIP-seq dataset shows that CarD is present when RNAP-σ^A is also found, supporting a model that CarD is present at the promoters of both up- and downregulated genes. However, our data is also compatible with an alternative model in which CarD acts directly as a monotonic transcriptional activator and genes that appear to be transcriptionally "repressed" by CarD are expressed at lower levels in WT bacteria due to decreased RNAP occupancy at non-CarD-activated promoters. If this alternative model were true, then we would expect to find RNAP-σ^A/CarD binding sites overlapping with transcription start sites (TSSs) ascribed to the transcriptionally "activated" promoters but absent from TSSs ascribed to transcriptionally "repressed" promoters. To examine the overlap between CarD binding sites and CarD-regulated transcripts, we used our RNA-seq dataset to identify a list of *M. smegmatis* genes whose transcript abundance was likely directly responsive to altered CarD-mediated RP_o stabilization activity based on having opposite expression patterns in $CarD^{R25E}$ versus $CarD^{I27W}$. To avoid internal genes within operons, we focused our analysis on 2,917 *M. smegmatis* genes directly downstream of a primary TSS (146) and categorized them into one of four classes (**Table S3**). TSSs associated with genes that were significantly down-regulated ($P_{adj} < 0.05$) in $CarD^{R25E}$ and significantly up-regulated in $CarD^{I27W}$ were classified as 'Activated' by CarD (n=117) while TSSs associated with genes that were significantly up-regulated in $CarD^{R25E}$ and significantly down-regulated in $CarD^{I27W}$ were classified as 'Repressed' by CarD (n=153). TSSs associated with genes that were significantly differentially expressed in both $CarD^{R25E}$ and $CarD^{I27W}$ but in the same direction relative to wild-type were classified as 'Uncategorized' (n=222) because their expression profile does not reflect the divergent expression pattern expected between CarD mutants with opposing effects on RP_o-stabilization *in vitro*. Lastly, any TSSs that were not significantly differentially expressed in both $CarD^{R25E}$ and $CarD^{I27W}$ were categorized as 'Not Significant' (n=2425).

We re-analyzed our previous ChIP-seq dataset (25, 103) and identified 1857 unique CarD binding sites across two biological replicates (**Table S3**). To avoid broad binding regions that may represent multiple, overlapping CarD binding sites, we focused on 1796 CarD binding sites less than or equal to 1000 base pairs (bp) in width. Of these 1796 CarD binding sites, 1390 sites (77.4%) overlapped with at least one mapped TSS in *M. smegmatis* and 1129 sites (62.8%) overlapped with a primary TSS associated with a protein-encoding gene (146). We examined the overlap between CarD binding sites and TSSs that were significantly differentially expressed in both $CarD^{R25E}$ and $CarD^{I27W}$ and found that 57.9% (285/492) of these TSSs were associated with CarD binding (**Table 1**). Among the differentially expressed genes, 53.0% (62/117) of 'Activated' TSSs and 67.3% (103/153) of 'Repressed' TSSs overlapped with a CarD binding site (**Table 1**). Thus, CarD binding is associated with transcriptional regulation of *M. smegmatis* promoters *in vivo* but is not correlated with the direction of regulation. A similar analysis was performed in the α-proteobacterium *Caulobacter crescentus* to identify the direct regulon of CdnL (the *C. crescentus* homolog of CarD) (107). Like our results, CdnL localized to promoter regions of both genes that were up-regulated and genes that were down-regulated in a Δ*cdnL* strain, but a vast majority of differentially expressed genes were not associated with CdnL binding, suggesting a broader effect of indirect regulation in *C. crescentus*. Together, these data support the model that CarD is broadly localized to mycobacterial promoters through its interaction with RNAP but that the regulatory outcome of CarD activity is not determined by occupancy.

CarD directly activates transcription from the *Mtb* ribosomal RNA promoter *rrnA*P3

To test our model that the outcome of CarD's RP_o stabilizing activity on transcript production depends on the basal promoter kinetics, we used *in vitro* transcription methods to measure the direct effects of CarD on transcript production. Although several studies have

proposed that CarD activates transcription from the *Mtb* AP3 promoter based on *in vitro* three-nucleotide transcription assays (59, 101) and real-time fluorescence assays (28, 46) that report RP_o lifetime, full-length transcript production has never been directly measured. To assess full-length RNA production, we performed multi-round *in vitro* transcription assays by incubating recombinantly purified *Mtb*RNAP-σ^A holoenzyme with a linear DNA fragment containing the *Mtb* AP3 promoter (from -39 to +4 with respect to the TSS) driving transcription of a 164 nucleotide RNA product. The addition of a saturating concentration of WT CarD (25:1 molar ratio CarD:RNAP (46)) activated transcription from the AP3 promoter ~8-fold compared to reactions with no factor added (**Fig. 2A**). To investigate how CarD's RP_o-stabilizing activity relates to transcriptional activation, we repeated the multi-round *in vitro* transcription assays with CarD mutants impaired in their ability to stabilize RP_o *in vitro* ($CarD^{R25E}$, $CarD^{R47E}$, $CarD^{K125A}$, and $CarD^{K125E}$) (46, 100). All four of the CarD mutants activated transcription from AP3 compared to reactions with no factor, but the degree of activation by each mutant was reduced compared to WT CarD (**Fig. 2A**), suggesting that CarD's RP_o-stabilizing activity underlies its ability to activate transcription from AP3. In addition, the degree to which each CarD mutant attenuated transcript production correlated with how severe the impact was on the CarD macromolecular interactions with RNAP (97) or DNA (100). In contrast, $CarD^{I27W}$, which has increased affinity for RNAP and is able to stabilize RP_o at lower concentrations than $CarD^{WT}$ (101), activated transcription from AP3 to a greater degree than $CarD^{WT}$ at concentrations below where $CarD^{WT}$ is saturating (5:1 molar ratio CarD:RNAP) (**Fig. 2B**), further demonstrating the association between CarD's RP_o-stabilizing activity and activation of transcript production. Collectively, these results demonstrate that CarD activates full-length RNA production *in vitro* from AP3 and this transcription activation is dependent on the RP_o-stabilizing activity of CarD.

Additional promoter DNA-RNAP interactions increase basal RP_o stability and push CarD towards transcriptional repression

To test the hypothesis that the degree of transcriptional activation by CarD is inversely correlated with the basal RP_o stability of a promoter, we explored CarD's direct regulatory effect on transcription from a set of promoters with varying levels of basal RP_o stability. Transcription initiation kinetics and RP_o lifetime are highly dependent on promoter DNA sequence (34). In the RP_o intermediate, the promoter DNA makes multiple sequence-specific contacts with regions of the RNAP holoenzyme to stabilize the transcription bubble (32, 34, 45). The *Mtb* RNAP-σ^A holoenzyme and WT AP3 ($AP3_{WT}$) promoter form a relatively unstable RP_o (46, 59) that is stabilized by CarD to lead to activation of transcription *in vitro* **(Fig. 2)**. We, therefore, used the AP3 promoter sequence as a starting point to generate four additional promoter templates ($AP3_{EcoExt}$, $AP3_{MycoExt}$, $AP3_{Discr}$, and $AP3_{Stable}$) with higher levels of basal RP_o stability by making targeted sequence mutations that would add or optimize predicted DNA-RNAP interactions in RP_o (**Fig. 3A**). $AP3_{WT}$ contains near consensus sequence motifs in the -35 and -10 elements (60), which are highly conserved promoter elements that interact with σ region 4 and 2, respectively (51, 53, 54), so we did not target these regions in our study. In $AP3_{EcoExt}$, we mutated the base at position -14 to a G to introduce a $T_{-15}G_{-14}$ motif that represents an extended -10 element that was first identified in *E. coli* (147). In addition to the classical *E. coli*-like extended -10 motif, many mycobacterial promoters instead contain a G at position -13 that is associated with promoter strength and RP_o formation in DNase I footprinting studies (61). Thus, we also generated $AP3_{MycoExt}$, which is mutated to include a G_{-13} upstream of the -10 element. Both G_{-14} and G_{-13} are positioned to interact with a conserved glutamic acid residue in σ^A region 3.0 in the mycobacterial RP_o (27, 28, 61). $AP3_{Discr}$ is mutated to introduce a $G_{-6}GGA_{-3}$ motif in the discriminator region

immediately downstream of the -10 hexamer that allows for optimal binding with σ^A region 1.2 (57, 148, 149). $AP3_{Stable}$ is mutated to include the mutations made in $AP3_{EcoExt}$ and $AP3_{Discr}$ as well as a deletion of a T at position -17 to reduce the length of the spacer region between the -35 and -10 hexamers from 18-bp in $AP3_{WT}$ to 17-bp. A spacer length of 17-bp allows for optimal interactions of the -35 and -10 hexamers with σ^A (150).

To measure the basal RP_o stability of RNAP-σ^A and the AP3 promoter variants, we performed *in vitro* three-nucleotide transcription initiation assays (59, 101) in the absence of CarD by incubating *Mtb*RNAP-σ^A holoenzyme with linear promoter DNA fragments in the presence of a GpU dinucleotide and UTP. In these reactions, the RNAP-σ^A holoenzyme can synthesize a three nucleotide 'GUU' RNA transcript but cannot undergo promoter escape, allowing us to assess relative RP_o lifetimes by using the amount of three nucleotide product as a proxy. We found that all the promoter variants with additional predicted DNA-RNAP contacts exhibited higher basal levels of RP_o stability compared to $AP3_{WT}$ (**Fig. 3B**). The most stable variant $AP3_{Stable}$ displayed 8-fold higher basal RP_o stability relative to $AP3_{WT}$. To quantify the effect of CarD on RP_o stability from these promoter variants, we also performed three-nucleotide transcription assays in the presence of WT CarD protein (**Fig. 3B**). On $AP3_{WT}$, CarD increased the amount of three nucelotide product by roughly 4-fold over reactions with no factor. As the basal RP_o stability of promoter variants increased, the degree of RP_o stabilization by CarD decreased to the point that on $AP3_{Stable}$ the addition of CarD resulted in no detectable difference in the amount of three nucleotide product.

Having established a set of promoters with different basal RP_o stability levels that range over nearly one order of magnitude, we performed multi-round *in vitro* transcription reactions using these AP3 promoter variants in the presence or absence of CarD to investigate the

relationship between the basal RP_o stability of a promoter and transcriptional regulation by CarD (**Fig. 3C**). We discovered that across the AP3 promoter variants, basal RP_o stability positively correlated with full length transcript production in the absence of CarD but negatively correlated with transcriptional activation by CarD. Indeed, the two promoters with the highest levels of basal RP_o stability ($AP3_{EcoExt}$ and $AP3_{Stable}$) were transcriptionally repressed by CarD, consistent with the predictions of our model and providing the first *in vitro* evidence of direct transcription repression by *Mtb* CarD.

Basal RP_o stability and CarD regulatory outcome are influenced by discriminator region guanosine + cytosine base pair frequency

In addition to forming direct interactions with the polymerase, promoter DNA sequences can also influence RP_o stability by affecting the chemical properties of the DNA molecule. For example, guanosine + cytosine base pairs in the discriminator region impose a kinetic barrier to DNA untwisting and unwinding during the formation of the transcription bubble due to their greater base-pairing and base-stacking stability compared adenosine + thymine base pairs (151, 152). Discriminator guanosine + cytosine base pair frequency (G+C%) is inversely correlated with RP_o stability (153) and has been shown to be a determinant of transcription control by DksA/(p)ppGpp (95, 154). To determine if changing the RP_o stability by modifying the G+C% of the discriminator affects the outcome of CarD activity on transcript production, we generated a set of AP3 promoter variants ($AP3_{Discr1}$ – $AP3_{Discr5}$) in which the discriminator region G+C% is titrated from 100% ($AP3_{Discr1}$) to 16.7% ($AP3_{Discr5}$) (**Fig. 4A**). We observed a negative correlation between discriminator G+C% and basal RP_o stability as measured by three-nucleotide transcription assays (**Fig. 4B**). CarD increased three nucleotide RNA production from all promoter variants tested, but

the magnitude of RP_o stabilization by CarD displayed a negative correlation with basal RP_o stability across AP3 variants as the discriminator G+C% was titrated (**Fig. 4B**).

Discriminator G+C% of the AP3 variants was also negatively correlated with basal transcript production in the multi-round transcription assay and the magnitude of transcription activation by CarD decreased as discriminator G+C% decreased (**Fig. 4C**). On the AP3 variant with the lowest discriminator G+C% and highest basal RP_o stability ($AP3_{Discr5}$), CarD decreased transcript production, further supporting that promoters with high basal RP_o stability can be transcriptionally repressed by CarD. Collectively, our experiments show that promoter sequence motifs that increase basal RP_o stability decrease the magnitude of transcriptional activation by CarD and can lead to transcriptional repression in the most stable RP_o contexts.

Promoter sequences that form more stable RP_o are associated with transcription repression by CarD *in vitro* and *in vivo*

We show that base substitutions in the spacer region, extended -10 region (**Fig. 3**), and discriminator (**Fig. 4**) can affect full-length transcript production and the direction of CarD regulation. To directly examine whether differences in relative RP_o stability could explain the outcomes in transcript production and CarD regulation we performed a linear regression analysis across all of our promoter templates (**Fig. 5**). For this analysis, the relative RP_o stability and relative transcription strength of each promoter variant was normalized to $AP3_{WT}$. In the absence of CarD, the rate of full-length transcript production shows a roughly linear positive correlation with the relative RP_o stability (**Fig. 5A**). In contrast, the $\log_2$ ratio of transcript production in multi-round transcription reactions +/- CarD shows a roughly linear inverse correlation with increasing RP_o stability, with the most stable promoter variants ($AP3_{EcoExt}$, $AP3_{Stable}$, and $AP3_{Discr5}$) being

transcriptionally repressed by CarD (**Fig. 5B**). The robust relationship across multiple promoter variants suggests that RP_o stability is a fundamental determinant of full-length transcript production and CarD regulatory outcome. Collectively, our experiments illustrate a relationship between RP_o stability, transcription strength, and CarD regulation and demonstrate that transcription factors like CarD can discriminate promoters based on their basal kinetic features to potentiate bidirectional outcomes in transcription regulation via a single kinetic mechanism.

Through our *in vitro* transcription experiments, we have identified multiple promoter sequence motifs associated with high RP_o stability *in vitro*. If our model is generally applicable to transcription from mycobacterial promoters throughout the genome, then we would expect to find an association between DNA sequence motifs associated with RP_o stability and transcriptional repression by CarD. To interrogate this prediction, we examined the prevalence of a consensus extended -10 motif ($T_{-15}G_{-14}N_{-13}$) and discriminator GC% in promoters that were differentially expressed in *our Mtb* and *M. smegmatis* RNA-seq datasets (**Table S5**). Since all of our *in vitro* experiments were performed in the context of a *Mtb*RNAP-σ^A holoenzyme, we limited our bioinformatic analysis to promoters containing a $A_{-11}NNNT_{-7}$ motif representing the consensus σ^A -10 element (60, 132, 136, 146), which comprised 90.5% (1609/1778) and 82.5% (2511/3043) of the primary TSSs in *Mtb* (132) and *M. smegmatis* (146), respectively. Indeed, in *Mtb*, promoters that were predicted to be repressed by CarD based on our RNA-seq data were over-enriched for extended -10 elements while promoters predicted to be activated by CarD were under-enriched for extended -10 elements relative to the genome-wide proportion of this feature (**Fig. S4A**). A similar trend was true of the proportion of promoters containing extended -10 elements in *M. smegmatis*, but the difference in proportions between CarD regulated promoters and the genome-wide distribution was not statistically significant (**Fig. S4B**). In both species, promoters that were

predicted to be repressed by CarD contained significantly more GC-rich discriminator regions than promoters predicted to be activated by CarD (**Fig. S4C-D**). The association of stable RP_o DNA sequence signatures with genes that are inferred to be repressed by CarD *in vivo* support that the regulatory mechanisms that we demonstrate *in vitro* could be relevant to gene expression *in vivo*.

DNA topology can influence the regulatory outcome of CarD activity

In mycobacteria, CarD transcript levels increase in response to double-stranded DNA breaks and genotoxic stress (63), suggesting that the dynamics of CarD regulation may be important for responding to these environmental cues. DNA breaks in the chromosome can relieve local regions of DNA supercoiling. The supercoiling state of promoters is tightly connected to transcriptional activity *in vivo*, as positive or negative supercoiling can inhibit or enhance RP_o formation, respectively (155, 156). Thus, we sought to test the relationship between promoter topology and CarD regulation. We generated a set of templates with identical DNA sequence but varied molecular topology by cloning the $AP3_{WT}$ promoter into a negatively supercoiled plasmid and incubating the plasmid with either a single-cutting endonuclease to produce a linear "cut" DNA molecule, a nicking endonuclease to produce a circular "nicked" DNA molecular, or with no enzyme to maintain a supercoiled control (mock treated) (**Fig. 6A**). We performed *in vitro* three-nucleotide transcription assays using the topologically distinct DNA templates and found that negative supercoiling contributes to a ~6-fold increase in basal RP_o stability compared to a linear "cut" DNA template containing the same promoter sequence (**Fig. 6B**). In addition, the "nicked" DNA template exhibited a similar basal RP_o stability to the "cut" DNA template, indicating that the higher RP_o stability observed in the "mock" template is a result of supercoiling and not the circular shape of the molecule. The addition of CarD decreased the amount of three nucleotide transcript produced with the supercoiled "mock" DNA template. This result could

indicate that CarD inhibits progression from RP_o towards an initial transcribing complex intermediate (RP_{itc}) that synthesizes the three nucleotide product quantified in these assays (75). The basal RP_o stabilities of the "cut", "nicked", and "mock" $AP3_{WT}$ DNA templates correlated with the basal transcriptional activity of the promoter, where promoter templates with high basal RP_o stability also showed high levels of basal transcript production (**Fig. 6C**). Furthermore, CarD activated transcription from the "cut" and "nicked" DNA templates but repressed transcription from the supercoiled "mock" DNA template, which has a higher basal RP_o stability relative to the "cut" and "nicked" molecules. These data demonstrate a single promoter DNA sequence can exhibit varying levels of basal RP_o stability based on DNA supercoiling, and this supercoiling-dependent change in RP_o stability can change the regulatory outcome of CarD on transcription. While the DNA sequence of a given promoter is constant within the genome, the topology of the DNA molecule can change over the lifetime of the cell. Thus, our findings reveal an additional layer of complexity in CarD's regulatory mechanism and could help explain how CarD expression *in vivo* could lead to differential gene expression outcomes in different conditions.

Discussion

CarD is an essential transcriptional regulator in *Mtb* that affects the expression of over two-thirds of the genome (111) and whose normal function and expression are required for bacterial survival during various stresses and virulence in mice (63, 97, 100, 102). Numerous *in vitro* studies have shown that CarD stabilizes RP_o formed by the housekeeping *Mtb*RNAP-σ^A holoenzyme (46, 59, 96), leading to the early model that it functions as a general transcription activator. However, a subsequent RNA-seq study of *Mtb* strains encoding mutant alleles of CarD revealed a more complex scenario where CarD appears to differentially activate or repress transcription from

different promoters (111). In an effort to understand how CarD could affect gene expression in a promoter specific manner, we now provide experimental evidence for a relationship between RP_o stability and the outcome of CarD regulation that results in promoter specific effects of CarD activity. We find that the ratio of transcript production in multi-round transcription reactions +/- CarD shows a roughly linear inverse correlation with increasing RP_o stability, with the most stable promoter variants ($AP3_{EcoExt}$, $AP3_{Stable}$, and $AP3_{Discr5}$) being transcriptionally repressed by CarD. CarD's effect on mycobacterial transcription *in vivo* also reflects the observations from our *in vitro* experiments where promoters predicted to be repressed by CarD are associated with sequence features that correlate with high RP_o stability (extended -10 sequence motif, low GC% discriminator region) and promoters predicted to be activated by CarD are associated with an absence of these features. Collectively, these data support our model in which the specific outcome of CarD-mediated RP_o stabilization is dependent on the kinetic properties of a given promoter and not on sequence-specific binding, which could explain the observed differential gene expression effects in CarD mutants *in vivo*.

Our study deepens our understanding of mycobacterial transcription regulation and demonstrates how RNAP-binding factors like CarD add complexity to this process. The true relationship between promoter sequence and CarD regulation is likely more nuanced than the data presented in this study. Although the AP3 promoter variants we generated were designed to increase or decrease RP_o stability in a stepwise manner (i.e. $AP3_{Stable}$ is a combination of $AP3_{Discr}$ and $AP3_{ExoEct}$; G+C% is titrated one base at a time in $AP3_{Discr1}$-$AP3_{Discr5}$), the effects of each mutation are likely more complex. Minor base substitutions in a promoter sequence can result in large-scale allosteric effects on other RNAP-DNA interactions (157) and kinetic steps (158) outside of RP_o. Furthermore, our model was built on the idea that CarD uses a single kinetic

mechanism, but CarD's effects on specific transcription initiation rate constants may differ between promoters. For example, CarD contains a conserved tryptophan residue in its DBD that is positioned to interact with a T at the -12 position of the non-template DNA strand at the upstream fork of the bubble (96), and it has been hypothesized that this sequence-specific interaction acts as a "wedge" to prevent bubble collapse. In theory, on a promoter lacking T_{-12}, CarD's inhibitory effect on $k_{collapse}$ may be diminished relative to its effects on the rates of bubble opening and promoter escape, producing a unique kinetic mechanism that is biased towards repression. In our RNA-seq dataset, *Mtb* promoters that were predicted to be repressed by CarD were significantly enriched for non-T bases at the -12 position (111), lending some *in vivo* support for the prediction that this DNA sequence context biases CarD towards transcriptional repression.

Another region that we did not study but could affect CarD regulation is the initially transcribed sequence downstream of the TSS, which can affect the kinetics of promoter escape and RNAP pausing (49, 137, 159). Simplistically, CarD represses transcription from certain promoters by over-stabilizing RP_o and decreasing transcript flux by inhibiting promoter escape (127), leading to an accumulation of abortive transcripts (140). However, this model becomes more complicated when considering a branched pathway of transcription initiation (50), where a fraction of RNAP form moribund complexes that never undergo promoter escape. Kinetic studies using a fluorescent reporter showed that CarD increases the fraction of unescaped RNAP complexes (75), and we show that in some contexts CarD can inhibit the synthesis of a three nucleotide product from RP_o. These data suggest that CarD could affect steps of initial nucleotide incorporation prior to promoter escape and influence the fraction of RNAP complexes undergoing productive versus moribund transcription (49, 160).

We also find that this relationship between RP_o stability and CarD is not only influenced by promoter sequence, where promoters with identical DNA sequence can be differentially activated or repressed depending on their supercoiling status (**Fig. 6**). In our experiments, CarD directly activated transcription from the *Mtb* rRNA promoter AP3 on a linear DNA template but repressed transcription from AP3 on a negatively supercoiled template (**Fig. 6**), which is the predominant topological state of DNA in bacterial cells (155). On the surface, this seems to contradict CarD's role as a positive regulator of rRNA synthesis *in vivo* (100, 101, 107, 109). However, one possible explanation may be that CarD is required to maintain efficient transcription of highly transcribed genes, such as the rRNA operon, when they accumulate positive supercoils due to their high transcriptional activity (155). We propose that CarD may function to overcome the topologically self-limiting nature of rRNA transcription to promote rapid bacterial growth.

Beyond its role in specifically regulating rRNA synthesis, CarD also affects the transcription of hundreds of other *Mtb* genes *in vivo* (111), which could explain CarD's pleiotropic effects under different stresses. *Mtb* strains with altered CarD activity are also sensitized to various environmental stresses other than nutrient starvation including oxidative stress, genotoxic stress, and antibiotic treatment (63, 97, 100, 102), but it is still unclear what roles CarD plays under these conditions. The impact of topology and promoter context also implies that CarD may elicit different effects on gene expression in different environments. CarD's ability to interpret the kinetic properties of a promoter add modularity to the mycobacterial transcription response, because while DNA sequence is essentially constant over the lifetime of a bacterial cell its kinetic properties may be dynamic and responsive to environmental stimuli. *In vivo*, the supercoiling state of a promoter is constantly changing in response to the translocation of polymerases, the enzymatic action of topoisomerases, and DNA damage caused by antibiotics or other genotoxic stresses (155,

161). In addition to DNA supercoiling, other environmental factors such as intracellular NTP concentrations (162) and temperature (163) can influence the kinetic properties of a promoter without affecting DNA sequence. During pathogenesis, *Mtb* may encounter these environmental stimuli in various combinations. Furthermore, the expression of CarD is itself highly responsive to environmental signals including nutrient limitation (102) and DNA damage (63). Understanding how transcription factors like CarD interact with these environmental stresses may provide insight into how *Mtb* responds to the host environment and antibiotic treatment, making this an intriguing direction of future study.

Based on the results of this study, we propose that CarD belongs to a growing class of RNAP-binding transcription factors that include DksA/(p)ppGpp (92, 93), TraR and its phage-encoded homologs (164, 165), and the σ-subunit interacting transcription factors including the Actinobacteria-specific protein RbpA (62, 69, 70, 73, 74). Like CarD, these factors coordinate broad transcriptional programs in bacteria (28, 74, 139), highlighting the expanded regulatory range of these factors compared to classical transcription factors that are limited to promoters containing a specific binding motif. All of these global transcriptional regulators function by modulating the kinetics of transcription initiation, albeit via different mechanisms. Whereas CarD stabilizes RP_o, DksA/(p)ppGpp binds RNAP and destabilizes a kinetic intermediate preceding RP_o, resulting in transcriptional repression at ribosomal RNA promoters that form unstable RP_o and transcriptional activation at promoters of amino acid biosynthesis genes that form relatively stable RP_o (89, 92–94, 166). Although they exert opposite effects on initiation kinetics, CarD and DksA/(p)ppGpp share the ability to "read" the kinetic properties of a promoter to exert multiple regulatory outcomes on transcription. This study of CarD's regulatory mechanism demonstrates how kinetic context influences the activity of this class of RNAP-binding transcription factors and

reveals another layer in how bacteria coordinate broad gene expression in response to their environment.

Acknowledgements

We thank Drake Jensen, Ana Ruiz Manzano, and Eric Galburt for their helpful discussions and for providing the *E. coli* protein expression strain for *Mtb*RNAP-σ^A purification. We also thank John Errico, Helen Blaine, and Daved Fremont for their help in purification of RNA polymerase.

Tables

	Differentially Expressed	**Not Significant**	**Activated**	**Repressed**	**Uncategorized**	**TOTAL**
Overlapping with CarD binding site	**285** (57.9%)	**1089** (44.9%)	**62** (53.0%)	**103** (67.3%)	**120** (54.0%)	**1374** (47.1%)
Total # TSSs	492	2425	117	153	222	2917

Table 1. CarD binding is associated with both activated and repressed transcription start sites (TSSs). 'Differentially Expressed' TSSs are those TSSs associated with genes that were significantly differentially expressed ($p_{adj}<0.05$) in both $CarD^{R25E}$ and $CarD^{I27W}$ relative to $CarD^{WT}$. TSSs associated with genes that were not significantly differentially expressed in both mutant strains were categorized as 'Not Significant'. P-values from a hypergeometric test are listed. Differentially expressed genes were categorized as: 'Activated' if it was down-regulated in $CarD^{R25E}$ and up-regulated in $CarD^{I27W}$, 'Repressed' if it was up-regulated in $CarD^{R25E}$ and down-regulated in $CarD^{I27W}$, or 'Uncategorized' if it was differentially expressed in the same direction in both mutant strains.

Figures

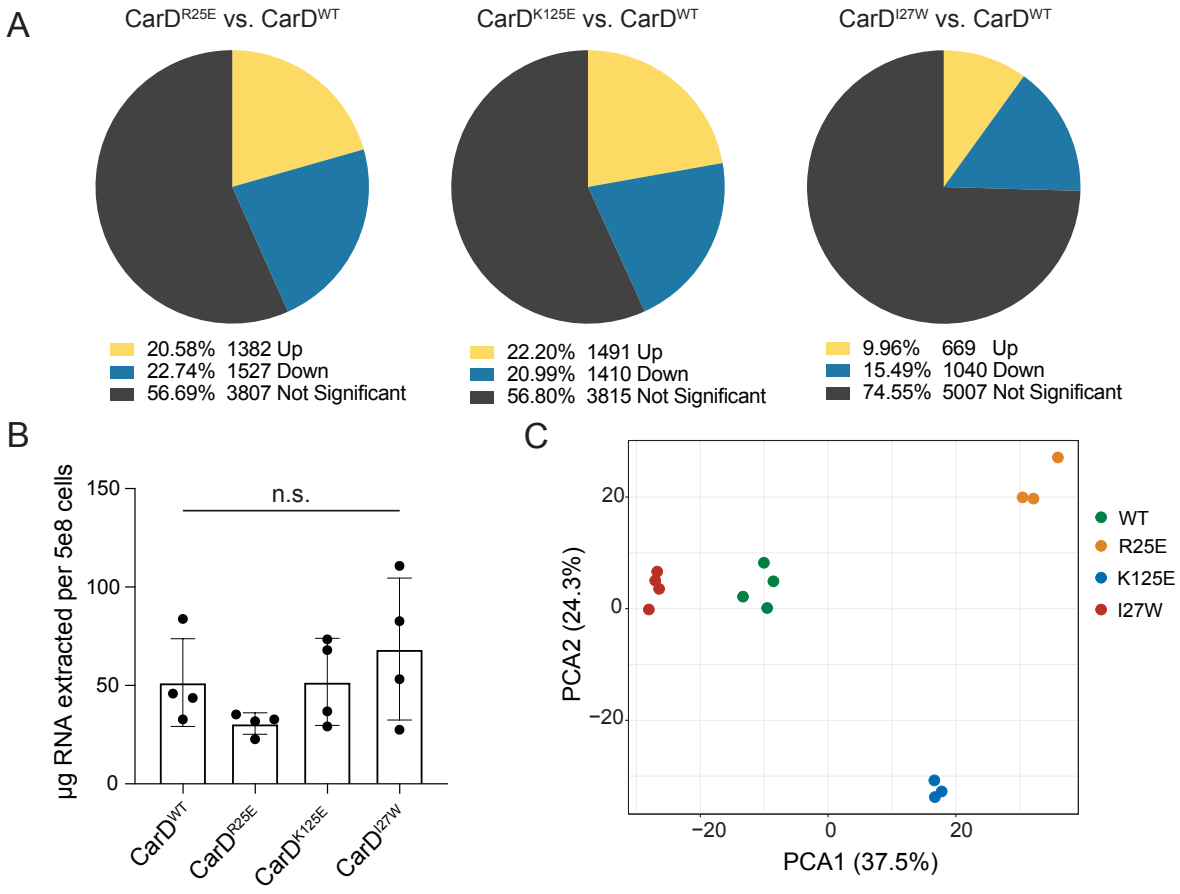


Figure 1. *M. smegmatis* strains encoding point mutants of CarD display broad changes in transcript expression. (A) Pie charts displaying the percentage of *M. smegmatis* coding genes that were significantly differentially expressed ($p_{adj} < 0.05$) in each CarD mutant strain relative to CarD^{WT}. (B) RNA content for *M. smegmatis* strains expressing different alleles of CarD calculated from total RNA weight harvested from four biological replicates divided by estimated number of cells collected. Each bar represents the mean ± standard deviation (SD). Group means were compared using a one-way ANOVA and determined to be not significantly different (p = 0.228). (C) Principal component analysis of RNA sequencing samples based on read counts of 6,716 *M. smegmatis* MC^{2}155 coding genes. The first two principal components (PC1 and PC2),

which account for 37.5% and 24.3% of the variance, respectively, define the x- and y-axis, respectively.

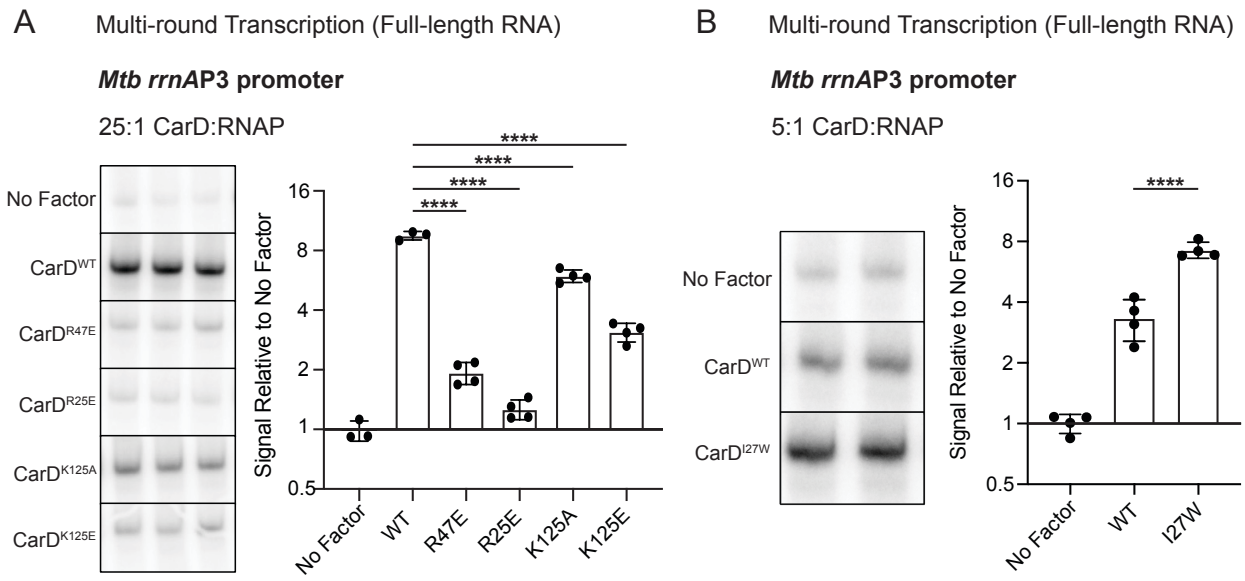


Figure 2. CarD activates transcription from the *Mtb* ribosomal RNA promoter AP3, and mutations to either the RNA polymerase (RNAP) interaction domain (RID) or DNA-binding domain (DBD) impair this activity *in vitro*. (A) Representative gel images from multi-round *in vitro* transcription reactions of *Mtb*RNAP-σ^A holoenzyme on linear DNA templates encoding AP3 with either no factor, wild-type CarD ($CarD^{WT}$), one of two RID mutants ($CarD^{R47E}$ or $CarD^{R25E}$), or one of two DBD mutants ($CarD^{K125A}$ or $CarD^{K125E}$). In all reactions with factor, CarD is added at a 25:1 molar ratio to RNAP holoenzyme. The bar graph displays the mean transcript signal intensity relative to 'No Factor' ± standard deviation (SD). N=3-4 independent reactions for each condition. (B) Representative gel images from multi-round *in vitro* transcription reactions of *Mtb*RNAP-σ^A holoenzyme on AP3 with either no factor, $CarD^{WT}$, or a RID mutant with higher affinity for RNAP ($CarD^{I27W}$). In all reactions with factor, CarD is added at a sub-saturating concentration of 5:1 molar ratio to RNAP holoenzyme. The bar graph displays the mean transcript signal intensity relative to 'No Factor' ± SD. N=4 independent reactions for each condition. (A and B) Mean fold-change values were compared using a one-

way ANOVA followed by post-hoc Dunnett's tests comparing the mean of each mutant CarD allele to $CarD^{WT}$; **** = $p<0.0001$. The full ANOVA results are listed in **Table S4**.

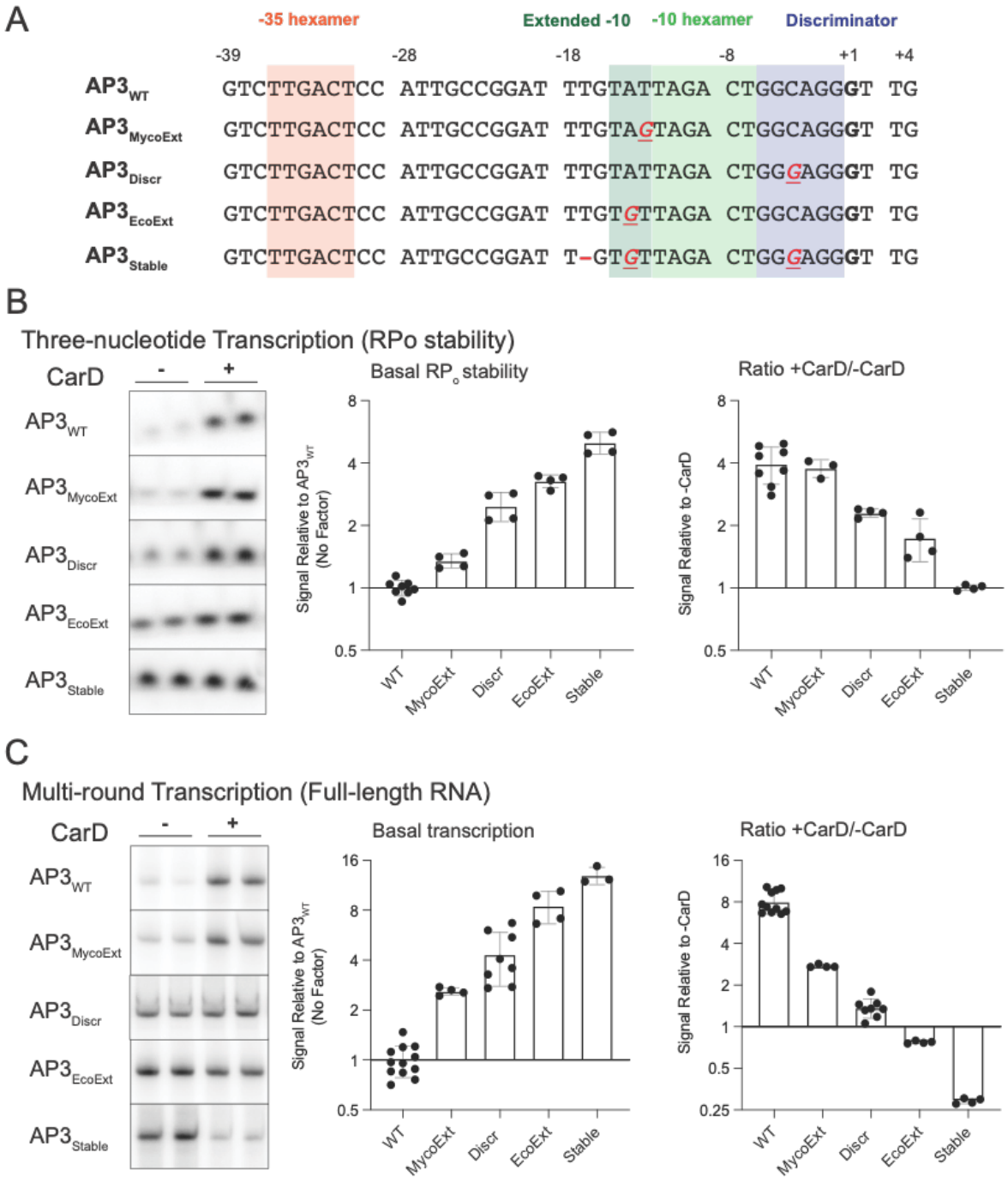


Figure 3. Promoter sequences that introduce additional interactions between promoter DNA and RNAP in the open complex increase basal RP_o stability and shift the regulatory outcome of CarD towards transcriptional repression. (A) Promoter sequences of the wild-type *Mtb rrn*AP3

promoter ($AP3_{WT}$) and four variants with sequence mutations that add predicted interactions between promoter DNA and RNAP in RP_o. Sequences from the -39 to +4 position relative to the transcription start site (+1, bolded) are shown. In the non-WT sequences, DNA bases that are altered from the WT sequence are underlined and colored red. A "—" indicates that a base was deleted. (B) Representative gels showing [^{32}P]-labeled three-nucleotide transcription products formed by *Mtb*RNAP-σ^A from linear DNA templates encoding $AP3_{WT}$ or one of the four AP3 variants either in the absence or presence of CarD. (C) Representative gels from multiround transcription assays showing 164 nucleotide [^{32}P]-labeled RNA transcripts produced by *Mtb*RNAP-σ^A from linear DNA templates encoding $AP3_{WT}$ or one of the four AP3 variants either in the absence or presence of CarD. (B and C) Bar graphs display (left) the mean basal signal intensity relative to $AP3_{WT}$ ± standard deviation (SD) and (right) the mean ratio of signal intensity +CarD/-CarD for each promoter ± SD Group means were compared by one-way ANOVA $p<0.0001$. The full results of pairwise comparisons are listed in **Table S4**.

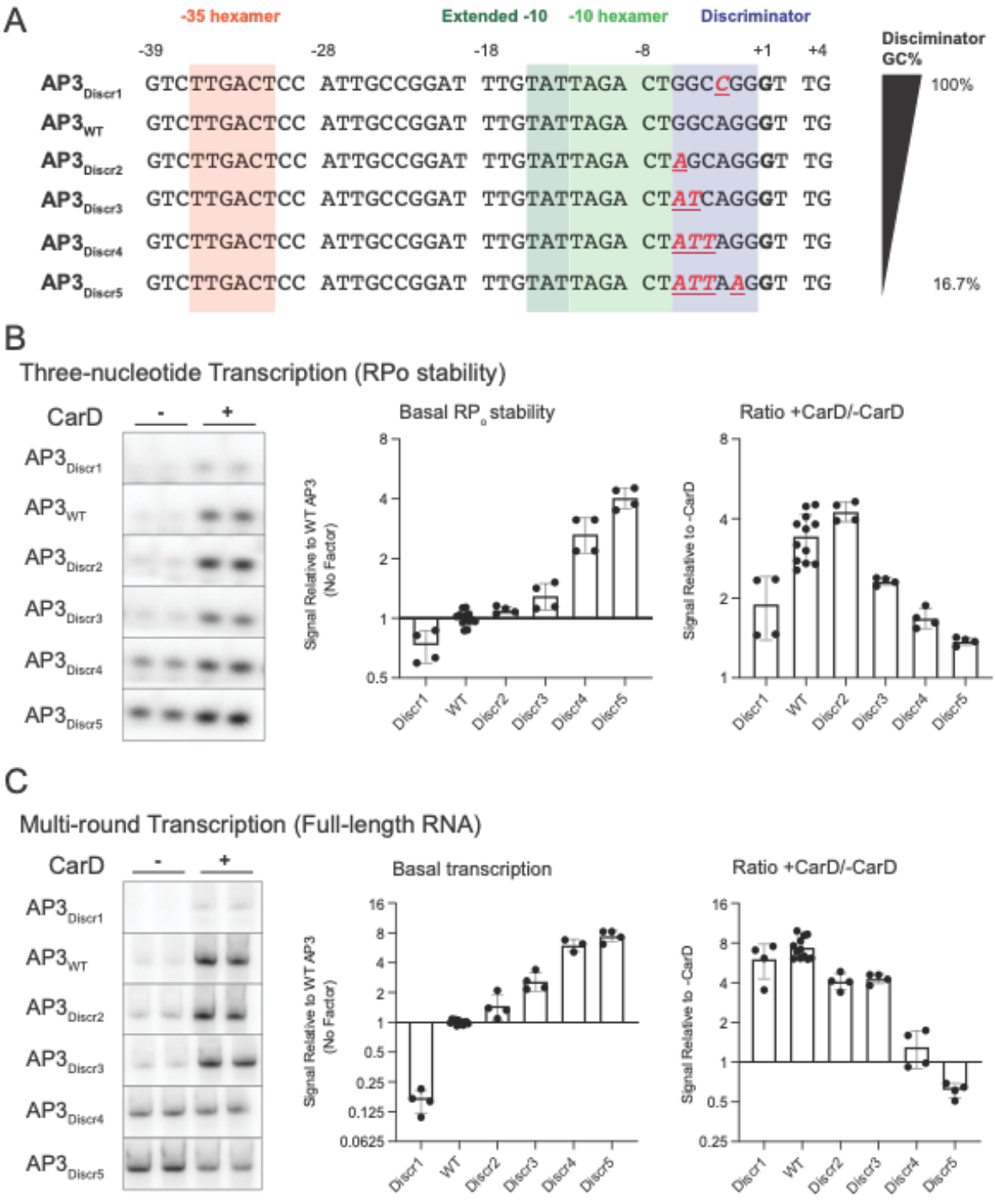


Figure 4. Discriminator GC% negatively correlates with basal RP_o stability and influences CarD regulatory outcome. (A) Promoter sequences of the wild-type *Mtb rrnA*P3 promoter ($AP3_{WT}$) and

five variants with sequence mutations that either increase or decrease the percentage of G or C bases in the discriminator. Sequences from the -39 to +4 position relative to the transcription start site (+1, bolded) are shown. In the non-WT sequences, DNA bases that are altered from the WT sequence are underlined and colored red. (B) Representative gels showing [^{32}P]-labeled three-nucleotide transcription products formed by *Mtb*RNAP-σ^A from linear DNA templates encoding $AP3_{WT}$ or one of the four AP3 variants either in the absence or presence of CarD. (C) Representative gels from multi-round transcription assays showing 164 nucleotide [^{32}P]-labeled RNA transcripts produced by *Mtb*RNAP-σ^A from linear DNA templates encoding $AP3_{WT}$ or one of the four AP3 variants either in the absence or presence of CarD. (B and C) Bar graphs display (left) the mean basal signal intensity (in the absence of CarD) relative to $AP3_{WT}$ ± standard deviation (SD) and (right) the mean ratio of signal intensity +CarD/-CarD for each promoter ± SD Group means were compared by one-way ANOVA $p<0.0001$. The full results of pairwise comparisons are listed in **Table S4**.

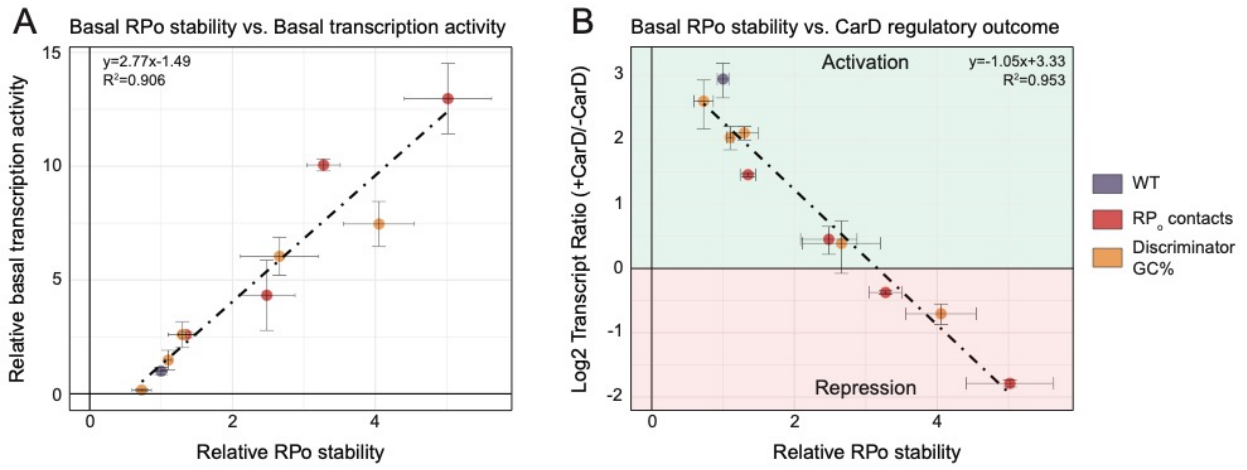


Figure 5. The basal RP_o stability of a promoter is positively correlated with basal transcription activity but negatively correlated with transcription activation by CarD. (A) Dot plot showing the relationship between the basal RP_o stability of AP3 promoter variants relative to $AP3_{WT}$ on the x-axis versus basal transcription activity relative to $AP3_{WT}$ on the y-axis. Each point represents a variant of the AP3 promoter and is colored based on whether it represents the WT promoter, a sequence with mutations that affect RNAP-DNA interactions in RP_o (RP_o contacts), or a sequence with mutations that affect the discriminator region GC% (Discriminator GC%). The position of each point represents the mean values from at least N=4 experiments and the error bars represent standard deviations. The dashed line and text represent the results of a linear regression analysis. (B) Dot plot showing the relationship between the basal RP_o stability of AP3 promoter variants versus the $\log_2$ ratio of transcript production in reactions ±CarD on the y-axis. Positive 'Log_2 Transcript Ratio' values indicate transcription activation while negative values indicate transcriptional repression.

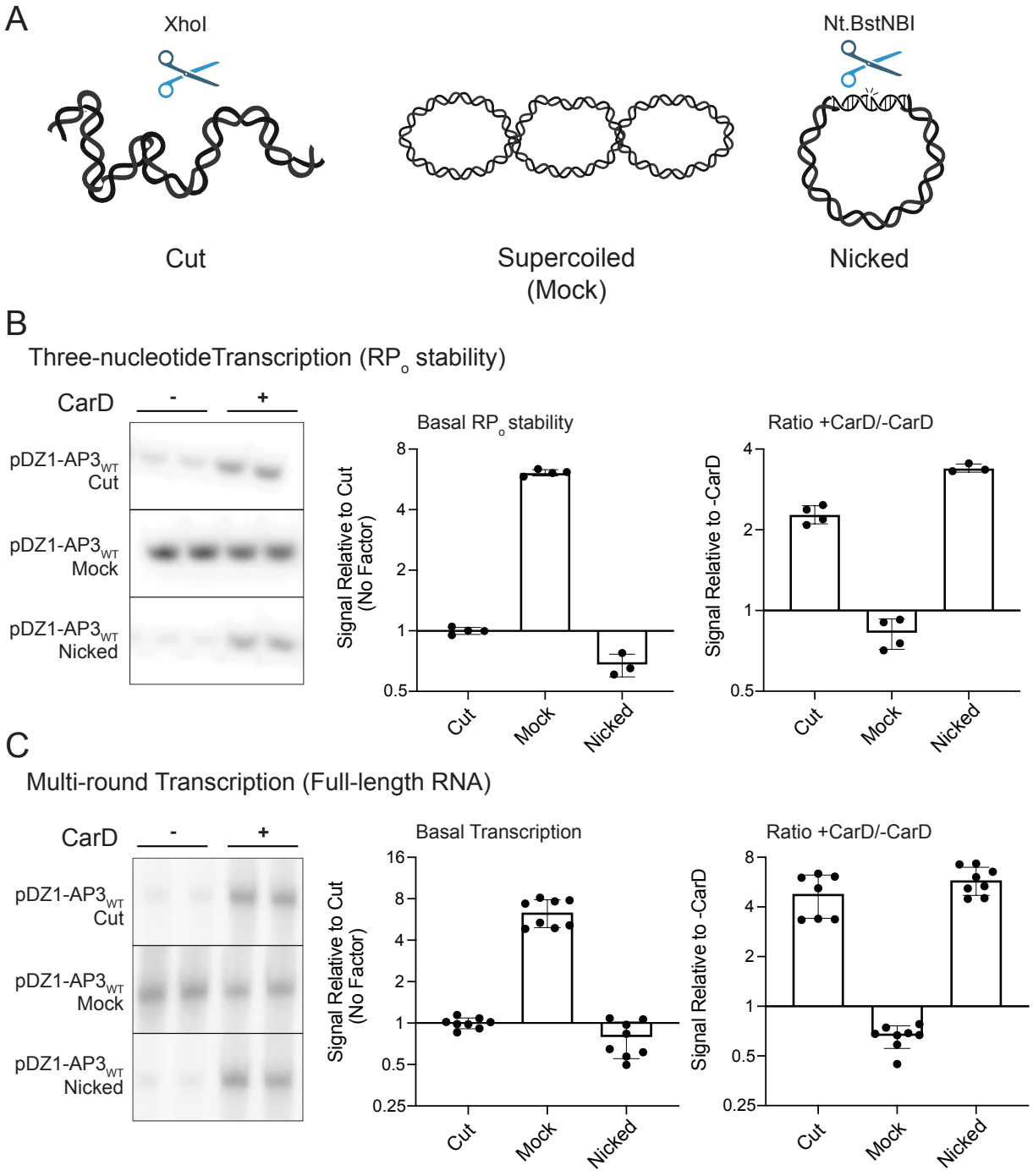


Figure 6. DNA topology affects the RP_o stability of promoters and can alter the regulatory outcome of CarD. (A) Schematic of DNA templates used for *in vitro* transcription reactions. All templates transcribe an identical ~100-nucleotide RNA product from the wild-type *Mtb rrn*AP3

promoter ($AP3_{WT}$). Each DNA template originates from a negatively supercoiled plasmid that was treated with no enzyme (Mock), a nicking endonuclease (Nicked), or a single-cutting restriction endonuclease (Cut). (B) Representative gels showing [^{32}P]-labeled three-nucleotide transcription products formed by *Mtb*RNAP-σ^A from each DNA template either in the absence or presence of CarD. (C) Representative gels showing full-length [^{32}P]-labeled RNA transcripts produced by *Mtb*RNAP-σ^A from each DNA template in either the absence or presence of CarD. (B and C) Bar graphs display (left) the mean basal signal intensity relative to the linear "Cut" DNA template ± standard deviation (SD) and (right) the mean ratio of signal intensity +CarD/-CarD for each promoter ± SD. Group means were compared by one-way ANOVA $p<0.0001$. The full results of pairwise comparisons are listed in **Table S4**.

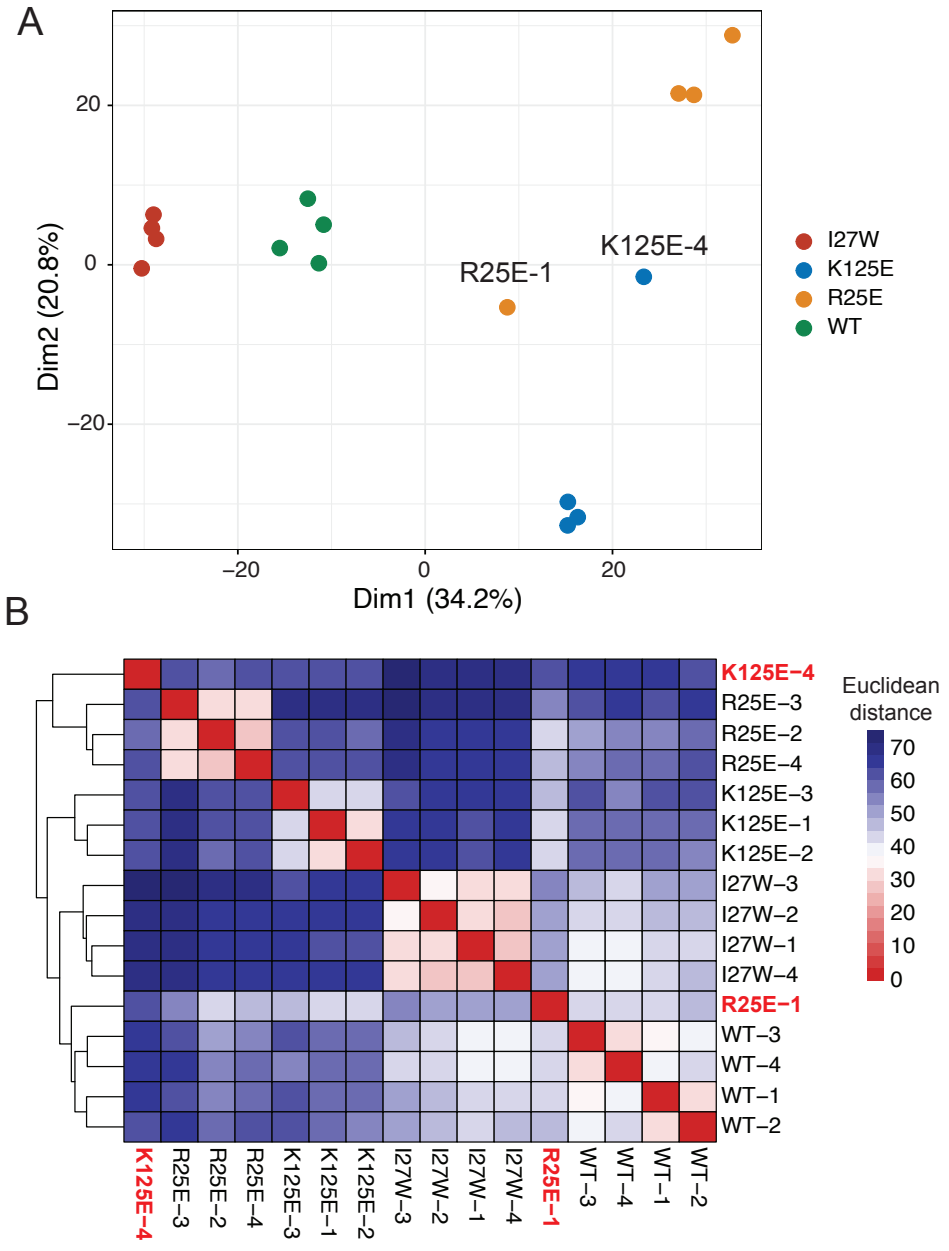


Figure S1. Two replicates from our RNA-sequencing (RNA-seq) experiment were removed from downstream analysis following outlier detection. (A) Principal component analysis of RNA-seq samples based on read counts of 6,716 *M. smegmatis* MC²155 coding genes. The points are colored by sample genotype. Samples that were removed as outliers are labeled with their sample ID. Both outlier samples were >1.4 standard deviations away from the mean on both

of the first two principal components. (B) Clustering of RNA-seq samples based on *M. smegmatis* transcript expression. The heatmap is colored based on the Euclidean distance between pairs of RNA-seq replicates, calculated based on read counts of *M. smegmatis* MC^2155 coding genes. Outlier samples are colored in red text and cluster with sequencing replicates from different genotypes than their own.

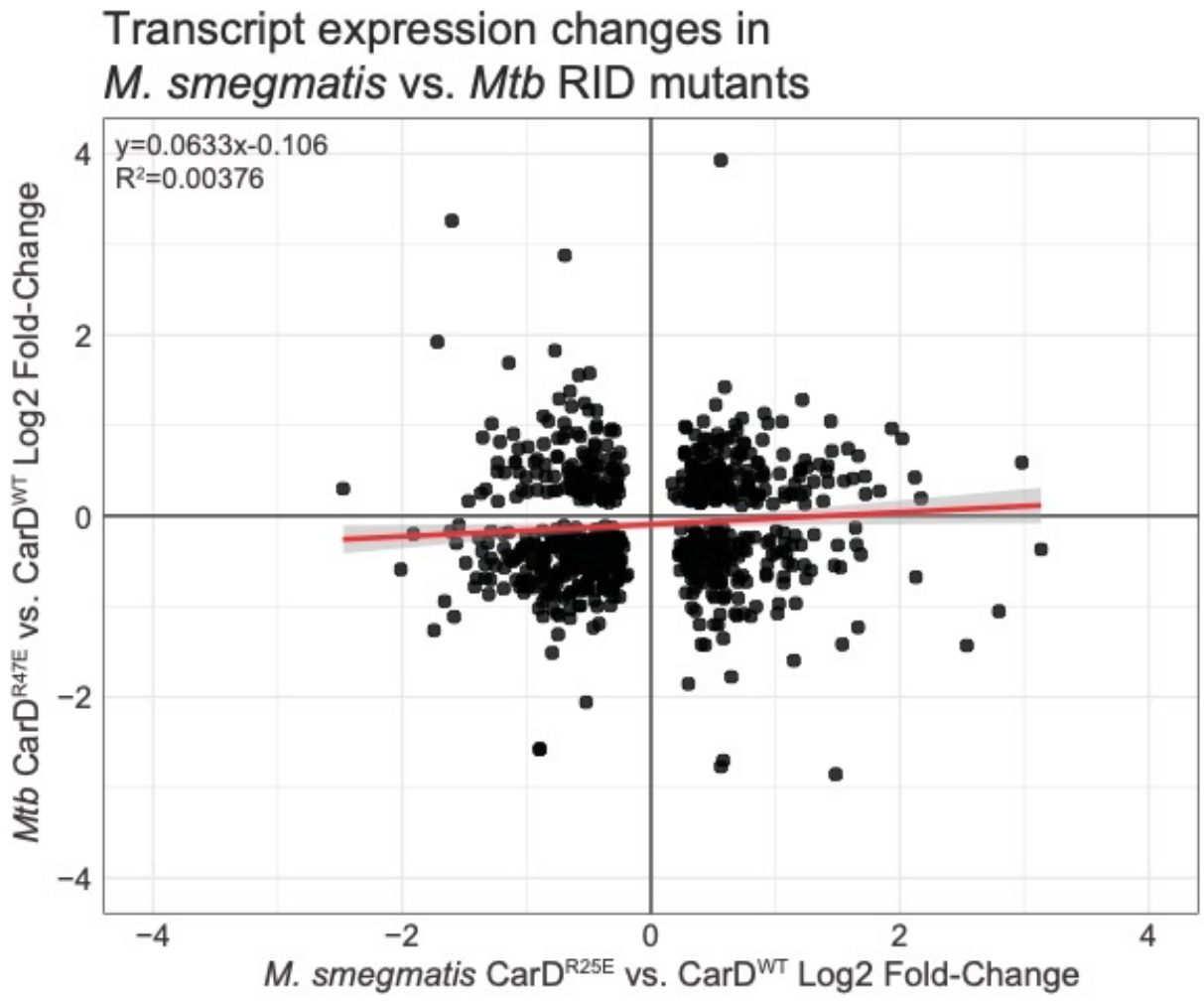


Figure S2. Comparison of transcript expression changes in CarD RNA polymerase interaction domain (RID) mutants in *M. smegmatis* versus *Mtb* (CarD^{R25E} and CarD^{R47E}, respectively) relative to strains expressing wild-type CarD (CarD^{WT}). Points represent 757 homologous protein-encoding genes that were identified by amino acid sequence identity and significantly differentially expressed in both CarD^{R25E} and CarD^{R47E} (**Table S2**). The red line is a linear regression line, and the shaded area around the line represents a 95% confidence interval.

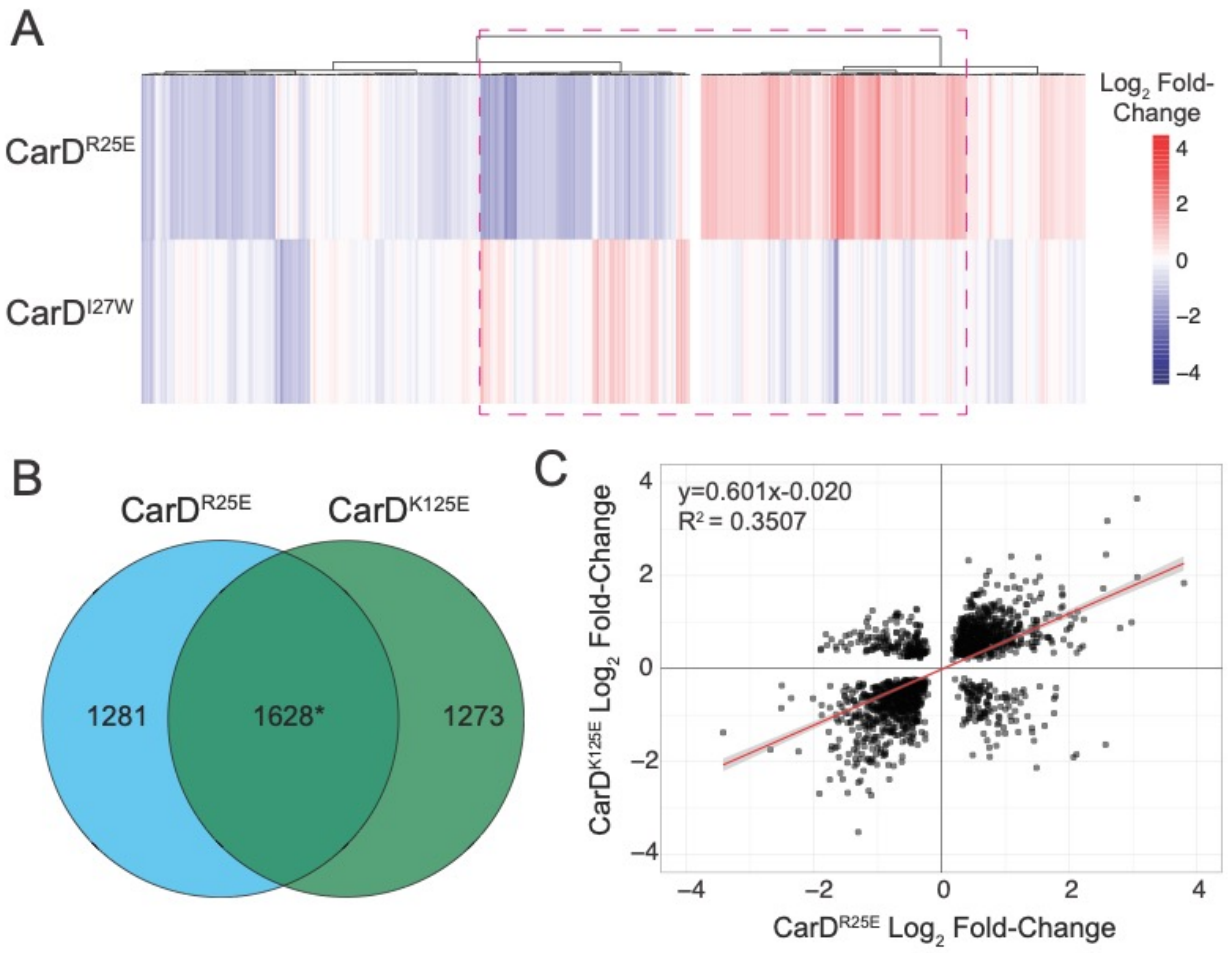
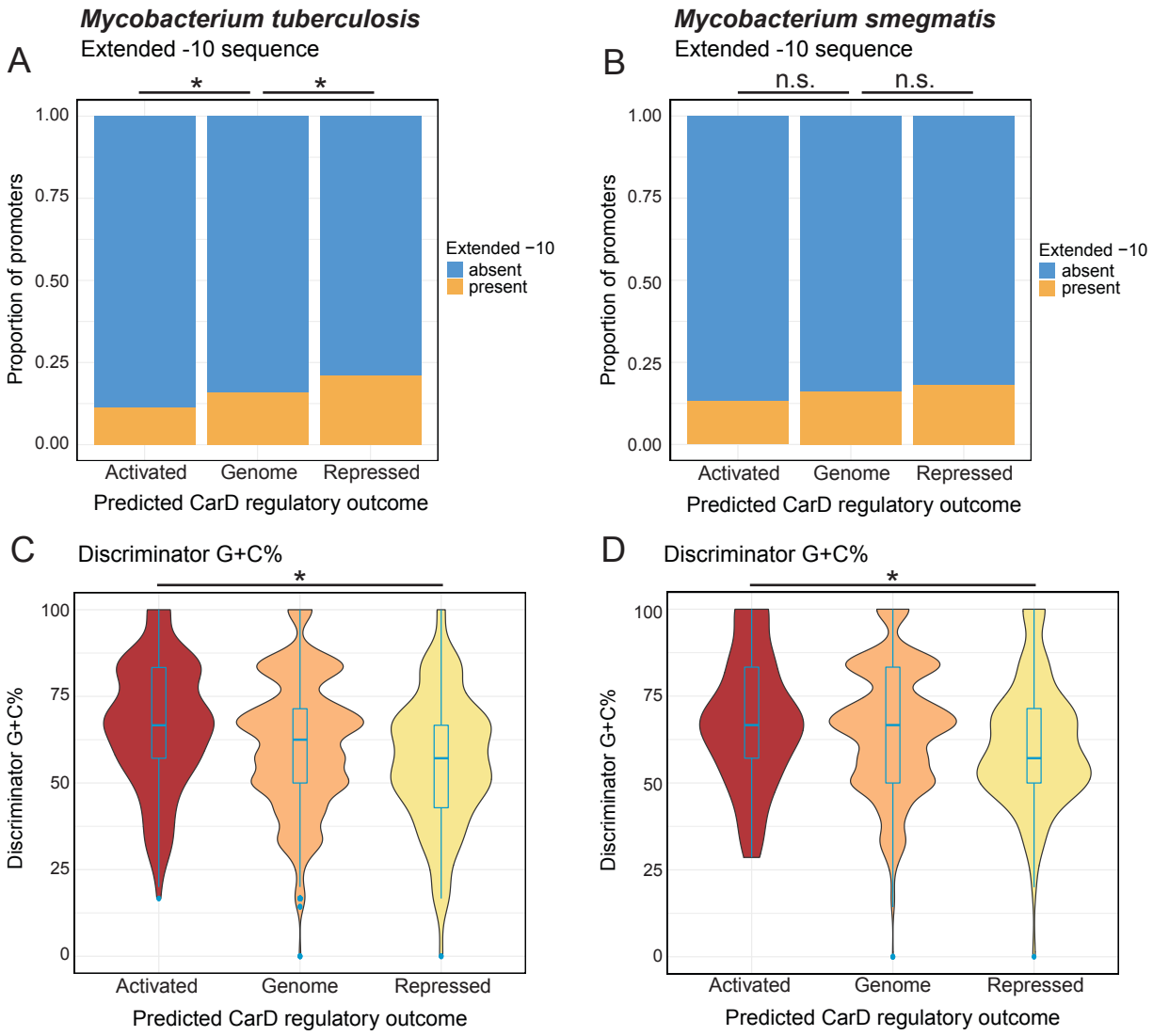


Figure S3. *M. smegmatis* strains expressing CarD RID mutants show largely opposite transcriptomic phenotypes, but there is little correlation in the transcriptomic changes between CarD^{R25E} and CarD^{K125E}. (A) Hierarchical clustering of 1,186 *M. smegmatis* coding genes that were significantly differentially expressed (p_{adj} <0.05) greater than 2-fold in at least one CarD mutant genotype. Genes were clustered using Ward's method based on the $\log_2$ fold-change in expression in each of the 3 CarD mutant strains relative to CarD^{WT}. 158/254 genes differentially expressed greater than 2-fold in CarD^{R25E} are differentially expressed in the opposite direction in CarD^{I27W}. The most highly differentially expressed genes in CarD^{R25E} (highlighted by the dashed magenta box) are differentially expressed in an opposite pattern in CarD^{I27W}. (B) Venn diagram

displaying the overlap in the lists of *M. smegmatis* genes significantly differentially expressed (p_{adj}<0.05) in CarD^{R25E} versus CarD^{K125E}. *The overlap is significant (p<0.001) based on a hypergeometric test. (C) Scatter plot displaying transcript expression changes for CarD^{K125E} and CarD^{R25E} on the y- and x-axes, respectively. The red line is a linear regression line, and the shaded area around the line represents a 95% confidence interval.



Supplemental Figure S4. Promoter sequence characteristics of *Mtb* and *M. smegmatis* promoters classified by their predicted CarD regulatory outcome determined based on RNA sequencing data. 'Genome' denotes the overall proportion of features across all TSSs in the genome. *Mtb* promoters were classified as 'Activated' if it was down-regulated in CarD^{R47E} and up-regulated in CarD^{I27W} or 'Repressed if it was up-regulated in CarD^{R47E} and down-regulated in CarD^{I27W}. *M. smegmatis* promoters were classified as 'Activated' if it was down-regulated in

CarD^{R25E} and up-regulated in CarD^{I27W} or 'Repressed if it was up-regulated in CarD^{R25E} and down-regulated in CarD^{I27W} (A-B) Bar plots displaying the proportion of promoters containing a consensus extended -10 sequence motif ($T_{-15}GN_{-13}$) in either *Mtb* (A) or *M. smegmatis* (B). Enrichment was tested using a hypergeometric test comparing the proportions of each group to the overall proportion of extended -10 promoters among promoters containing a SigA-like -10 sequence motif; * = $p<0.05$, n.s. = not significant. (C-D) Violin plots displaying the discriminator G+C nucleotide base percentage of promoters in *Mtb* (C) or *M. smegmatis* (D). Box and whisker plots displaying the median and interquartile range are drawn in blue. Mean discriminator G+C% values were compared using a Kruskall-Wallis rank sum test followed by post-hoc Dunn's tests for pairwise comparisons; * = $p<0.05$. The full results of our promoter analysis are listed in **Table S5**.

Chapter 4: Connecting *Mycobacterium tuberculosis* transcription factor networks to CarD regulation

Dennis X. Zhu

Introduction

The ability of a transcription factor to meaningfully contribute to bacterial stress response and survival depends on two characteristics: 1) the capacity of the transcription factor to alter its activity in response to a stimulus and 2) the capacity of the transcription factor to alter the expression of a specific regulon of genes. Compared to the numerous studies that have explored the kinetic mechanism of CarD in mycobacteria (25, 28, 46, 59, 64, 75, 96, 100, 101), relatively few studies have directly examined the regulation of CarD expression (102) or the phenotypic effects of its activity (63, 97, 100, 101).

Recently, a study from Li *et al*. elucidated a complex network of post-transcriptional regulation that governs intracellular CarD protein levels in *Mtb* and *M. smegmatis* (102). In their model, CarD is abundant during exponential growth under nutrient replete conditions and rapidly down-regulated upon nutrient limitation. This down-regulation occurs at two levels: 1) translation of *carD* mRNA is blocked by the SigF-dependent transcription of an antisense RNA (asRNA) encoded on the opposite strand of the *card* gene and 2) up-regulation of ClpC1, a starvation specific unfoldase that targets CarD protein for Clp protease-mediated degradation (167). Together, these two systems induce a sharp decline in CarD protein levels upon carbon, nitrogen, or phosphorus starvation or upon entry into stationary phase. This pattern of CarD expression matches studies in *Rhodobacter sphaeroides*, where CarD protein levels are similarly high during exponential growth and depleted during stationary phase (110). The existence of an asRNA or Clp-mediated degradation have not been demonstrated in *R. sphaeroides*, but *in vitro* transcription experiments show that *Rsp*CarD can inhibit transcription from its own promoter (140). However, as the authors mention, this mechanism alone is not sufficient to produce the expression patterns of CarD *in vivo*.

CarD mutants in mycobacteria display wide-ranging phenotypes including sensitivity to killing by oxidative stress, nutrient starvation, DNA damage, and various antibiotics (63, 97, 100, 101). The breadth of these phenotypes is unsurprising given CarD's broad effects on the mycobacterial transcriptome (111). Due to the essentiality of CarD in mycobacteria and the size of its regulon, it is difficult to tease apart which of these phenotypes are a consequence of direct CarD regulation and which are due to indirect effects of CarD depletion or mutation. Many of the regulatory effects of CarD in *M. smegmatis* are likely due to indirect regulation, as only 57.9% of genes differentially expressed in a CarD mutant strain are associated with a CarD binding event in our ChIP-seq dataset (**Chapter 3, Table 1**).

In this chapter, I present measurements of CarD protein levels during different phases of *Mtb* growth that support the current model of CarD regulation in the field. In addition, I present an exploratory analysis of the transcription factor networks that are perturbed in the *Mtb* CarD mutants. My analysis reveals that the expression of numerous transcription factors is dysregulated in the CarD mutant strains, but further examination of their regulons shows that the pattern of dysregulation in the CarD mutant strain is often not consistent with transcription factor activity. My analysis identifies the regulons two *Mtb* transcription factors GlnR (Rv0818) and ArgR (Rv1657) as pathways that are significantly correlated with CarD activity.

Experimental Procedures

Bacterial growth, lysate collection, and Western blotting

Wild-type *Mtb* Erdman strain cultures were grown in Middlebrook 7H9 media supplemented with 10% albumin/dextrose/catalase (ADC) and 0.02% tyloxapol. Bacterial cell lysate collection was performed by pelleting bacteria and resuspending cells in NP-40 lysis buffer (10mM sodium

phosphate pH 8.0, 150mM sodium chloride, 2mM EDTA, 0.1% NP-40, 1mM DTT, 1x Roche cOmplete protease inhibitors) to an effective OD_{600} = 30. The cell suspension was then bead beat, mixed 1:1 (v/v) with 2x Laemmeli buffer and heated for 20 minutes at 90° C to sterilize bacteria.

Western blotting was performed by running the samples on a NuPAGE 4-12% Bis-Tris acrylamide gel and transferring to Amersham Protran Western Blotting nitrocellulose membrane. Samples were blotted with a monoclonal mouse α-CarD antibody (clone 10F05; Memorial Sloan Kettering Cancer Center Monoclonal Antibody Core Facility) and a monoclonal mouse α-RpoB antibody (clone 8RB13; BioLegend). Visualization was performed by blotting with a secondary horseradish peroxidase conjugated goat α-mouse IgG antibody. Densitometry was performed using Image Lab software (Bio-Rad).

Results

CarD protein levels in *Mtb* decline during stationary phase and can be rescued by supplementation with fresh media

The growth of bacterial populations in liquid culture can be modeled into four different phases: lag phase, log phase, stationary phase, and death phase. Bacterial replication reaches its maximum rate during log phase, when nutrient availability in the culture is not yet limiting and cells are continuously synthesizing new material. During stationary phase, the population's growth rate levels off, and cells begin to shut down expensive cellular processes such as the synthesis of stable RNAs (88). Prior *in vitro* transcription studies of CarD implicate it as positive regulator ribosomal RNA (rRNA) biosynthesis (25, 100). Therefore, we hypothesized that CarD may show growth phase dependent expression within *Mtb*. Specifically, we hypothesized that CarD would be expressed at the highest levels during log phase and decline in stationary phase.

To test our hypothesis, I cultured wild-type *Mtb* in liquid culture and collected samples of bacterial cell lysate at different points in its growth phases to measure CarD protein levels by Western blotting. *Mtb* was cultured in the presence of a non-metabolizable detergent Tyloxapol to prevent cellular clumping, which could decrease the accuracy of cell density measurements. I collected cell lysates from 2 biological replicates at five different growth phases: early log phase (OD_{600} = 0.5-0.6), mid log phase (OD_{600} = 0.8-1.2), late log phase (OD_{600} = 1.5-2.0), early stationary phase (OD_{600} = 3.0-4.0), and late stationary phase (OD_{600} > 6.0). To test whether the expression of CarD protein could recover from growth-phase induced changes, I back-diluted cells from late stationary phase cultures into fresh media, allowed the cultures to double over 16 hours, and collected 'recovered' cell lysate samples. Western blotting shows that CarD is maximally expressed during early log phase and continuously declines as cell density rises (**Figure 1**). CarD expression is faint but still detectable even in late stationary *Mtb* cultures with an OD_{600} > 6.0. Back-dilution of bacteria into fresh media results in a recovery of CarD protein expression to a level similar to cultures prior to back-dilution (**Figure 1**), suggesting that stationary phase induced down-regulation of CarD protein levels is reversible. These data support the model that CarD expression is highly regulated by bacterial growth phase.

The ArgR and GlnR transcription factor regulons are dysregulated in CarD mutant *Mtb* strains

Improper CarD regulation has been implicated in sensitizing *Mtb* to multiple environmental stresses. However, the underlying transcriptional programs that explain why CarD function is necessary for bacterial viability during these stresses is unclear. In this section, I leverage the wealth of RNA-sequencing data that we generated to explore the transcription factor networks that are dysregulated in *Mtb* strains harboring CarD mutations.

Bacterial stress responses are most often controlled by the expression of a dedicated regulon under the control of a specific transcription regulator. Since the regulatory effects of CarD on the *Mtb* transcriptome are broad and CarD binds to promoters in a sequence-independent manner, I hypothesized that CarD regulates these *Mtb* stress response regulons indirectly by affecting the expression of upstream transcription factors. This hypothesis is supported by the fact that at least 42.1% of the differentially expressed genes in *M. smegmatis* mutant CarD strains relative to *Msm*CarD^{WT} are not associated with a CarD binding site, suggesting that these genes are targets of indirect regulation.

To examine the effect of CarD regulation on *Mtb* transcription factor networks, I began by examining the RNA transcript expression of transcription factors genes in *Mtb* strains expressing a mutant allele of CarD (R47E, K125A, I27F, I27W) relative to an *Mtb* strain expressing wild-type CarD (**Figure 2**). Specifically, I examined the expression of 13 sigma factor genes, 8 anti-sigma factor genes, 23 two-component system genes, and 183 *Mtb* transcription factor genes (168). For down-stream analysis I focused on transcription factors that fit into three criteria: 1) the transcription factor was significantly differentially expressed in all four CarD mutant strains; 2) the transcription factor is expressed in opposite directions in CarD^{R47E}/CarD^{K125A} versus CarD^{I27F}/CarD^{I27W}; 3) there exists at least one additional publication defining the regulon or function of the transcription factor in addition to the transcription factor overexpression dataset published by Rustad *et al*. (168) that was used to define my initial list of transcription factors. Fifteen transcription factor genes fulfilled these criteria: three sigma factors (*sigF*, *sigH*, and *sigM*), one anti-sigma factor (*rskA*; regulates SigK), one two-component system (SenX3/RegX3), and ten other transcription factors (*whiB3*, *mce2R*, *glnR*, Rv1353c, Rv1460, *argR*, *zur*, Rv2642, Rv3160c, and *whiB3*) (**Figure 2**, bolded).

The transcript expression of a transcription factor does not necessarily reflect a change in its activity, because many bacterial transcription factors are subject to post-transcription and post-translational regulation. If a given transcription factor is functionally activated or repressed by CarD, then the regulon of that transcription factor should be differentially expressed in a consistent manner in a CarD mutant background. Statistical analysis of transcription factor activity was performed based on a modified version of the regulon enrichment test (RET) described in Matern *et al*. 2018 (169). The RET used in this study essentially asks the question, do the genes in the regulon of X transcription factor behave in this condition as if the transcription factor is active. In this case, the condition that I am examining is the change in CarD activity in the $CarD^{R47E}$ *Mtb* strain compared to $CarD^{WT}$. To perform the RET, I classified all genes in our condition as either "up" or "down" in both our $CarD^{R47E}$ condition and "up" or "down" in response to a given transcription factor based on the *Mtb* transcription factor overexpression dataset published by Rustad *et al*. 2014 (168, 170). Since $CarD^{R47E}$ is a reduced function mutant, I categorized genes that were down-regulated in the mutant as "up" (i.e. activated by CarD) and genes that were up-regulated in the mutant as "down" (i.e. repressed by CarD). Following this classification scheme, I calculated an enrichment score (S) for each transcription factor using the following equation:

$$S = N_{Up,Up} + N_{Down,Down} - (N_{Up,Down} + N_{Down,Up})$$

where $N_{Up,Up}$ represents the number of genes that were classified as "up" in both the $CarD^{R47E}$ condition and in the transcription factor overexpression dataset. Genes that were present in the transcription factor regulon but not significantly differentially expressed in $CarD^{R47E}$ are classified as $N_{not\text{-}called}$. These genes do not affect the calculation of S but do contribute to the calculation of

the *p*-value. To normalize the enrichment score to the size of the regulon, I then calculate a mu-score (μ):

$$\mu = S/N_{regulon}$$

where $N_{regulon}$ represents the total number of genes in the regulon. The μ value ranges between -1.0 and +1.0, where a larger positive value means that in our unknown condition the regulon is expressed in the same direction as if the transcription factor is active and a larger negative value means that the regulon is being expressed in the opposite direction. Thus, in our specific case, a large positive μ score would suggest the regulon is indirectly activated by CarD while a large negative μ score would suggest that the regulon is indirectly repressed by CarD.

To calculate a *p*-value for the observed S, I performed a two-tailed hypothesis test based on 5000 Monte-Carlo simulations of random samples from a hypergeometric distribution with the same proportion of N_{Up}, N_{Down}, and $N_{not\text{-}called}$ genes as our CarD^{R47E} condition. For each sub-sample, I calculate a sample enrichment score (S'). The *p*-value of each transcription factor set is then calculated as:

$$p = P(|S| > |S'|)$$

Essentially, this significance test asks how often a randomly selected set of genes from the condition set would produce a regulon with an enrichment score with an absolute value higher than or equal to the enrichment score calculated for the transcription factor in our experimental condition.

RET analysis of the fifteen transcription factors identified above shows that only one transcription factor regulon, WhiB5, was significantly enriched ($p < 0.05$) in our RNA-seq dataset (**Table 1**). However, several other regulons displayed low *p*-values or large μ-scores, so I decided to expand my analysis to any regulons with a *p*-value less than or equal to 0.25, which added the

regulons for GlnR, ArgR, WhiB3, and SigK (**Table 1**). Out of the transcription factors with a *p*-value less than 0.25, two factors, GlnR and ArgR, had a |μ| value greater than 0.5, indicating a high degree of correlation between the expression pattern of genes in this regulon and CarD mutation. For both GlnR and ArgR, μ was negative, indicating that these two regulons are likely to be indirectly repressed by CarD. I examined the transcript expression of genes belonging to the SigK, ArgR, WhiB5, GlnR, and WhiB3 regulons in the $CarD^{R47E}$, $CarD^{WT}$, and $CarD^{I27W}$ *Mtb* RNA-seq samples (**Figure 3**) and found that the expression patterns for GlnR and ArgR matched the results of my RET analysis. The expression pattern of genes in the GlnR and ArgR regulons correlated highly with CarD genotype and were more highly expressed in the $CarD^{R47E}$ impaired function background and less highly expressed in the $CarD^{I27W}$ enhanced function background (**Figure 3**), consistent with the prediction that these regulons are indirectly repressed by CarD. The expression patterns of genes belonging to the SigK, WhiB3, and WhiB5 regulons were more varied, consistent with their low μ scores according to the RET analysis.

Discussion

My measurements of CarD protein levels in *Mtb* support the model of nutrient-dependent CarD regulation presented by Li *et al*. (102). The increase in CarD protein levels following back-dilution that I observed is similar to Li *et al*.'s data where supplementation with an additional carbon source could rescue CarD protein levels in stationary phase, suggesting that the underlying cause for CarD down-regulation in stationary phase is due to nutrient availability. The OD_{600} boundaries used to define log phase and stationary phase differ between our studies, but this can be explained by differences in culturing methods.

Based on the expression pattern of CarD in *Mtb*, I would hypothesize that it promotes the transcription of genes required for rapid growth and division while repressing transcription of genes involved in stationary phase survival and dormancy. Out of 208 *Mtb* transcription factor, sigma factor, and two-component system response regulator genes that I analyzed, only two transcription factors were significantly differentially expressed in all four CarD mutant *Mtb* strains sequenced in our RNA-seq study and showed significant expression changes in their regulon – GlnR and ArgR.

*Mtb*GlnR (Rv0818) is a homolog of the OmpR-family transcriptional factor GlnR that serves as the primary nitrogen-response regulator in *Streptomyces* and *M. smegmatis* (171, 172). Due to differences in their lifestyles, the nitrogen utilization pathways have diverged in *M. smegmatis* and *Mtb*. In *M. smegmatis*, GlnR is a positive regulator of genes involved in transport and assimilation of ammonium, which is the preferred nitrogen source of the bacteria and can be scavenged from their environment (172, 173). However, ammonium is not freely available to *Mtb* within host macrophages, and instead *Mtb* relies on metabolism of host amino acids and even reactive nitrate species (RNS) for nitrogen. *Mtb* metabolizes nitrate by first reducing nitrate to nitrate via the NarGHJI nitrate reductase complex and then reducing nitrate to ammonium using the nitrite reductase NirBD (174). GlnR binds to the promoters and activates the transcription of both the *narGHJI* and *nirBD* operons in *Mtb* (175). The nitrate and nitrite reductase systems of *Mtb* are essential for detoxification of RNS and survival in a hypoxic-induced *in vitro* dormancy model (174, 176). Our RNA-seq analysis shows that GlnR is significantly up-regulated in $CarD^{R47E}$/$CarD^{K125A}$ and down-regulated in $CarD^{I27F}$/$CarD^{I27W}$. In my RET analysis, the GlnR pathway had a negative μ score (-0.599), suggesting that expression of the regulon is negatively correlated with CarD regulation. Transcript expression of the *narGHJI* and *nirBD* operons is

highest in CarD^{R47E} and lowest in CarD^{I27W}, supporting the model that CarD negatively regulates nitrate and nitrite metabolism by repressing transcription of GlnR.

ArgR (Rv1657) is a homolog of phylogenetically conserved arginine repressor transcription factors that repress the transcription of arginine biosynthesis pathways when bound by L-arginine. Overexpression of ArgR in *Mtb* resulted in strong up-regulation of 7 out of 8 genes *argB-D, argF-H, and argJ* in the de novo arginine biosynthesis pathway (168), suggesting that in its apo- form ArgR activates transcription of this pathway. *Mtb* strains that lose function of this pathway are L-arginine auxotrophs, accumulate intracellular reactive oxygen species (ROS) and DNA damage, and are quickly cleared from mice. ArgR is significantly up-regulated in CarD^{R47E}/CarD^{K125A} and down-regulated in CarD^{I27F}/CarD^{I27W}, suggesting that CarD transcriptionally represses ArgR expression. In my RET analysis, the ArgR regulon had a μ score of -1.0 indicating that it is perfectly negatively correlated with CarD regulation. Thus, our RNA-seq dataset provides evidence that CarD is an indirect negative regulator of the *Mtb* de novo arginine biosynthesis pathway.

Three other transcription factor networks were significantly enriched in the CarD mutant RNA-seq dataset: SigK, WhiB3, and WhiB5. All three of these pathways had a low μ score (<0.5) in my RET analysis, indicating that although the regulon is being differentially expressed the direction of expression is not strongly indicative of transcription factor activation. These transcriptional patterns suggest that while these regulons may be responsive to CarD regulation, they may be separated by more than one degree of separation. For example, WhiB3 is part of the GlnR regulon (175, 177), so CarD may be influencing the expression of the WhiB3 regulon indirectly through two degrees of separation. The branches of indirect regulation may even occur through non-protein factors. For example, SigK, WhiB3, and WhiB5 are all sensors of redox

environment – WhiB3 and WhiB5 through their Fe-S cluster (83) and SigK through its redox-sensitive interaction with its anti-sigma factor RskA (178). In theory, defects in GlnR and ArgR expression would lead to the accumulation of RNS and ROS, respectively, within the bacteria, so the activities of SigK, WhiB3, and WhiB5 may be altered due to an indirect environmental change caused by CarD mutation. The possible regulatory interactions between just these five transcription factors highlight the difficulty in defining direct functional targets of CarD regulation in *Mtb*.

Tables

Gene ID	Name	*p*-value	μ score
Rv0022c	WhiB5	**0.046**	-0.378
Rv0491	RegX3	0.421	-0.364
Rv0586	Mce2R	0.742	0.167
Rv0818	GlnR	**0.164**	-0.599
Rv1353c		0.66	-0.281
Rv1460		0.488	0.250
Rv1657	ArgR	**0.158**	-1.00
Rv2359	Zur	0.663	-0.714
Rv2642		0.92	-0.667
Rv3160c		0.26	0.333
Rv3416	WhiB3	**0.214**	-0.431
Rv3286c	SigF	0.3	-0.431
Rv3223c	SigH	0.268	-0.429
Rv3911	SigM	0.525	-0.067
Rv0445c	SigK	**0.216**	0.100

Table 1. Results of regulon enrichment test (RET) analysis for fifteen *Mtb* transcription factors that were differentially expressed in all four CarD mutant strains relative to $CarD^{WT}$. The *p*-value represents the likelihood of observing the expression pattern of a given regulon in a randomly sampled gene set compared to our RNA-seq dataset. The μ score indicates the correlation between the direction of expression of genes in the regulon and how they were expressed in the CarD RNA-seq dataset. Transcription factors with a bolded *p*-value are visualized in **Figure 3**.

Figures

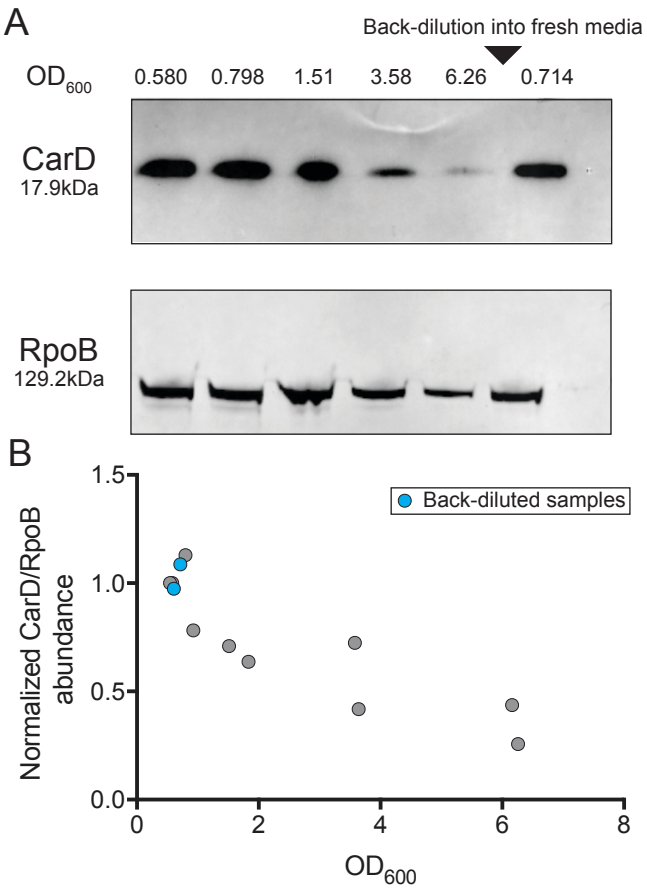


Figure 1. Western blot measurements of CarD protein levels during different growth phases of *Mtb*. (A) Representative Western blots of CarD and a loading control RpoB. (B) Quantification of relative CarD protein levels during different phases of growth. CarD levels were first normalized to RpoB abundance from the same sample and then normalized to the relative abundance of CarD at the first collection point ('early log phase') within each biological replicate (n = 2). Samples collected following back-dilution are colored in blue.

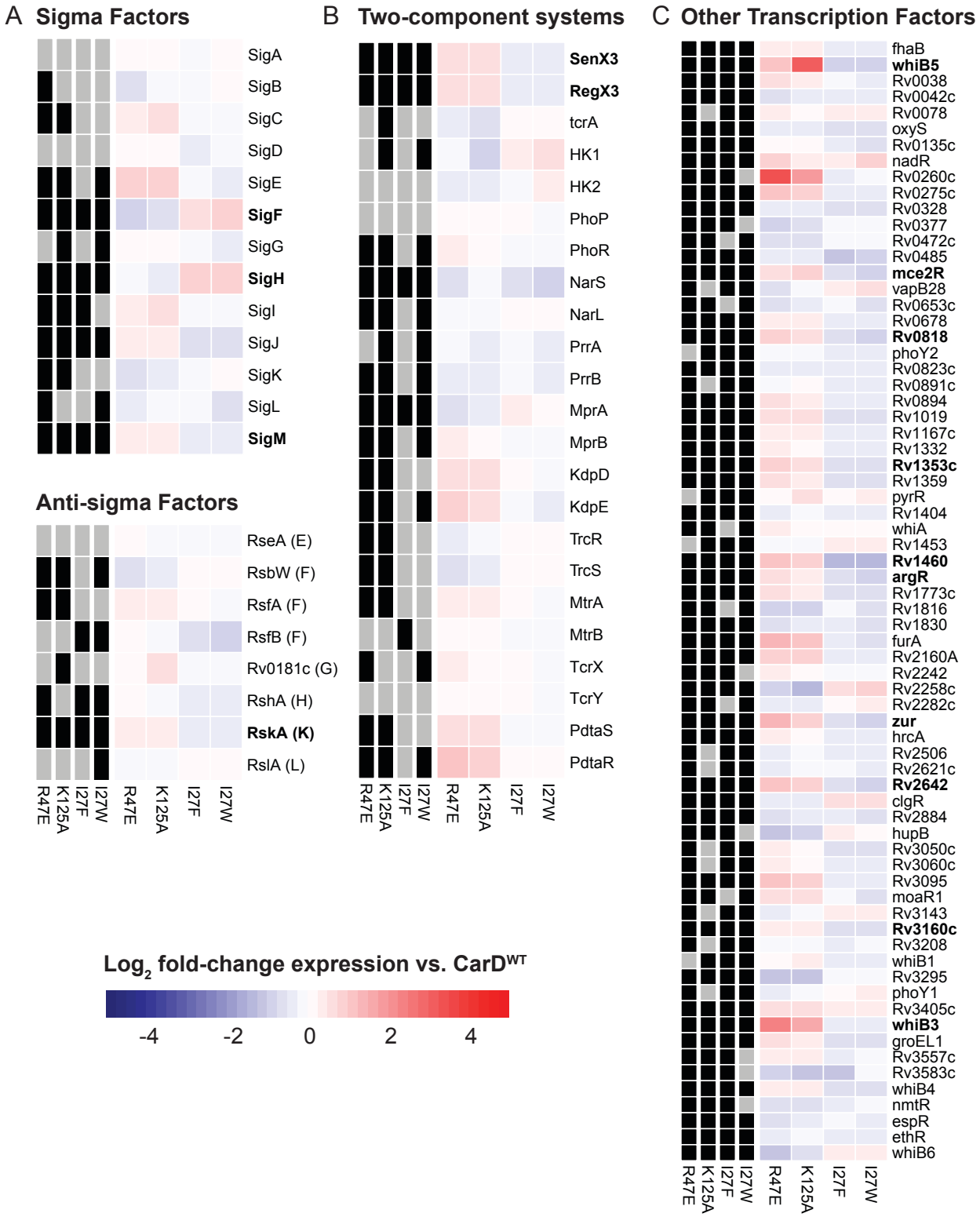

A Sigma Factors
SigA
SigB
SigC
SigD
SigE
SigF
SigG
SigH
SigI
SigJ
SigK
SigL
SigM
Anti-sigma Factors
RseA (E)
RsbW (F)
RsfA (F)
RsfB (F)
Rv0181c (G)
RshA (H)
RskA (K)
RslA (L)
R47E
K125A
I27F
I27W
B Two-component systems
SenX3
RegX3
tcrA
HK1
HK2
PhoP
PhoR
NarS
NarL
PrrA
PrrB
MprA
MprB
KdpD
KdpE
TrcR
TrcS
MtrA
MtrB
TcrX
TcrY
PdtaS
PdtaR
C Other Transcription Factors
fhaB
whiB5
Rv0038
Rv0042c
Rv0078
oxyS
Rv0135c
nadR
Rv0260c
Rv0275c
Rv0328
Rv0377
Rv0472c
Rv0485
mce2R
vapB28
Rv0653c
Rv0678
Rv0818
phoY2
Rv0823c
Rv0891c
Rv0894
Rv1019
Rv1167c
Rv1332
Rv1353c
Rv1359
pyrR
Rv1404
whiA
Rv1453
Rv1460
argR
Rv1773c
Rv1816
Rv1830
furA
Rv2160A
Rv2242
Rv2258c
Rv2282c
zur
hrcA
Rv2506
Rv2621c
Rv2642
clgR
Rv2884
hupB
Rv3050c
Rv3060c
Rv3095
moaR1
Rv3143
Rv3160c
Rv3208
whiB1
Rv3295
phoY1
Rv3405c
whiB3
groEL1
Rv3557c
Rv3583c
whiB4
nmtR
espR
ethR
whiB6
Log2 fold-change expression vs. CarDWT
-4
-2
0
2
4

Figure 2. Differential expression of *Mtb* transcription factors in *Mtb* CarD mutants relative to $CarD^{WT}$. Heatmaps show the log_2 fold-change in transcript expression for (A) *Mtb* sigma factor and anti-sigma factor genes, (B) *Mtb* two-component system genes, and (C) 70 *Mtb* transcription factors described in a transcription factor over-expression study published by Rustad *et al*. 2014. For visualization purposes, the list of 183 transcription factors originally published was filtered down to 70 genes that were significantly differentially expressed in at least 3 of the CarD mutant strains relative to $CarD^{WT}$. Genes that were used for downstream analysis are bolded.

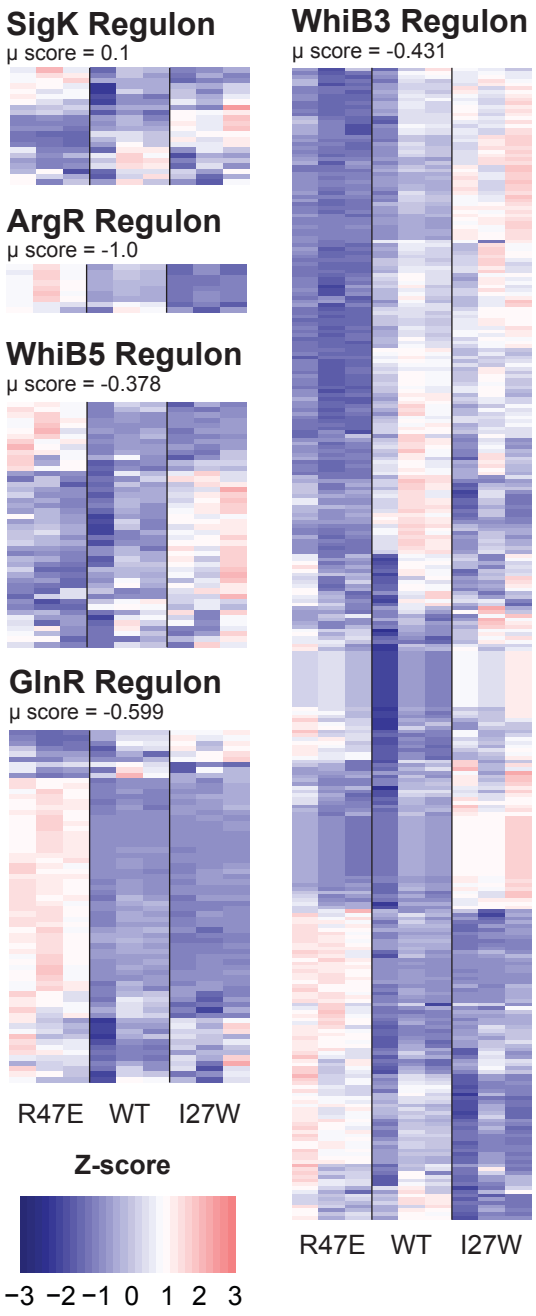


Figure 3. Transcript expression of genes belonging to the SigK, ArgR, WhiB5, GlnR, and WhiB3 transcription factor regulons in three biological replicates of CarD^{R47E}, CarD^{WT}, and CarD^{I27W} *Mtb* RNA-seq samples. The color of each cell represents the Z-score for the transcript expression of each gene across the nine samples displayed.

Chapter 5: Conclusions

Dennis Zhu

Major Findings

CarD is a broad, bi-directional transcription regulator in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

At the beginning of my PhD research project in 2017, CarD had been characterized as an essential transcription accessory protein required by the *Mycobacterium tuberculosis* (*Mtb*) RNA polymerase (RNAP) to form stable promoter open complexes (RP_o) during transcription initiation (25, 46, 59). CarD was known to accomplish this by binding directly to the RNAP β-subunit and interacting with promoter DNA to stabilize the open transcription bubble (25, 96, 100). Importantly, CarD's interaction with DNA is non-sequence specific and chromatin immunoprecipitation sequencing (ChIP-seq) studies showed that CarD localized ubiquitously to *M. smegmatis* promoters transcribed by the housekeeping σ^A RNAP holoenzyme (25, 98). Based on its ability to stabilize RP_o, the prevailing model of CarD's function in mycobacteria was as a general transcriptional activator. However, detailed kinetic studies of CarD's effect on transcription had only been performed on a handful of mycobacterial promoters, so CarD's role in regulating transcription throughout the mycobacterial genome had not yet been studied.

To characterize CarD's regulatory role at all mycobacterial promoters, I performed RNA-sequencing (RNA-seq) on *Mtb* and *M. smegmatis* strains encoding mutant alleles of CarD with altered activity (111). I found that in both *Mtb* and *M. smegmatis* over 2000 genes, representing over 50% of the protein-encoding genes in *Mtb* (**Chapter 2**, **Figure 1**) and over 40% of the protein-encoding genes in *M. smegmatis* (**Chapter 3**, **Figure 1**), were differentially expressed in CarD mutant strains compared to strains encoding a wild-type allele of CarD. Contrary to the model of CarD as a general transcription activator, CarD mutation resulted in roughly an equal number of up-regulated and down-regulated transcripts in both *Mtb* and *M. smegmatis* and did not result in a

global decrease in intracellular RNA levels (**Chapter 2**, **Figure S1**; **Chapter 3**, **Figure 1**). Our RNA-seq studies corroborate previous ChIP-seq studies that characterize CarD as a broad regulator of transcription throughout the mycobacterial genome. Furthermore, our RNA-seq analysis overturns the previous model of CarD as a general transcription activator and reveals a previously unappreciated role of CarD as a bi-directional transcription regulator.

CarD directly activates full-length RNA transcription from *Mtb* ribosomal RNA promoter *rrnA*P3

Prior to my thesis work, *in vitro* transcription studies of CarD's activity at mycobacterial ribosomal RNA (rRNA) promoters had concluded that CarD directly activates transcription from these promoters (25, 59, 101). However, the *in vitro* transcription methods used in these studies either tracked the formation of a three-nucleotide initial RNA product or only measured a single round of transcription. Therefore, these approaches do not account for CarD's inhibitory effect on promoter escape (75, 111). To overcome this experimental limitation, I used a multi-round *in vitro* transcription assay in which RNAP is incubated with a DNA template to initiate multiple rounds of transcription initiation, promoter escape, elongation, and termination. I performed multi-round *in vitro* transcription assays using *Mtb*RNAP-σ^A holoenzyme and a DNA template containing the *Mtb rrnA*P3 promoter in the presence or absence of CarD and found that wild-type (WT) CarD protein activates full-length transcript production roughly 7-fold from *rrnA*P3 (**Chapter 3**, **Figure 2**). This study was the first experiment to show that CarD directly activates transcription from a mycobacterial *rrnA* promoter even when accounting for CarD's inhibitory effect on promoter escape.

CarD's interactions with RNAP and DNA are both necessary for transcription activation and underlie its regulatory activity *in vivo*

Genetic studies using point mutations to disrupt CarD's interactions with DNA or RNAP showed that both of these interactions are necessary for the healthy growth of mycobacteria and the full virulence of *Mtb* in mice (97, 100). Furthermore, disrupting either of these interactions also impairs CarD's ability to stabilize RP_o *in vitro* (46, 100). Since I had already demonstrated that CarD could directly activate transcription from *rrnA*P3, I performed *in vitro* transcription experiments to dissect the roles of CarD's macromolecular interactions on its ability to activate transcription. I purified CarD proteins with mutations in its RNAP interaction domain (RID) (R25E and R47E) or its DNA-binding domain (DBD) (K125E and K125A) and performed multi-round *in vitro* transcription. All of the CarD mutants showed reduced transcription activation on *rrnA*P3 compared to WT (**Chapter 3**, **Figure 2**). The degree of transcription activation also correlated with the severity of the mutation in weakening CarD's interaction with either RNAP or DNA. Previous genetic studies also characterized mutations in CarD's RID (I27F and I27W) that increased its affinity for RNAP (101). These I27F and I27W CarD mutants increase the effective concentration of CarD and activate transcription from *rrnA*P3 at lower concentrations of protein compared to WT CarD (**Chapter 3**, **Figure 2**).

In my RNA-seq studies, I sequenced strains of *Mtb* encoding the CarD mutants R47E, K125A, or I27W ($CarD^{R47E}$, $CarD^{K125A}$, and *Mtb*$CarD^{I27W}$) and *M. smegmatis* strains encoding the CarD mutants R25E, K125E, or I27W ($CarD^{R25E}$, $CarD^{K125E}$, and *Msm*$CarD^{I27W}$). This experimental design allowed me to compare the effect of each of these mutations on CarD's regulatory role *in vivo*. In *Mtb*, $CarD^{R47E}$ and $CarD^{K125E}$ showed remarkably similar transcriptomic profiles and a roughly opposite transcriptomic signature compared to *Mtb*$CarD^{I27W}$ (**Chapter 2**, **Figure 3**).

Similarly, CarD^{R25E} and *Msm*CarD^{I27W} also showed opposite transcriptional profiles (**Chapter 3**, **Figure S3**). Collectively, these RNA-seq data indicate that CarD's RID and DBD interactions function together to potentiate CarD's *in vivo* regulatory activity and that CarD's *in vivo* effect on the mycobacterial transcriptome is correlated with its ability to bind RNAP and stabilize RP_o.

A majority of the promoters that are transcriptionally responsive to CarD mutation in *M. smegmatis* are directly bound by CarD

In *M. smegmatis*, CarD affects the transcription of nearly three thousand genes yet only binds to around 1800 distinct sites on the chromosome (103), meaning that many of the transcript abundance changes *in vivo* are due to indirect regulatory effects. However, I have also shown that CarD can directly activate transcription from promoters *in vitro*, so at least some portion of the transcriptional changes *in vivo* should be due to the direct activity of CarD at a promoter. The lack of a DNA binding motif for CarD's non-sequence specific DBD further hinders the identification of CarD's direct regulon. To identify the genes that are directly regulated by CarD, I cross-referenced our *M. smegmatis* RNA-seq dataset with the previously published ChIP-seq dataset (103) and found that 57.9% (285/492) of the differentially expressed genes associated with a mapped transcriptional start site (TSS) (146) were also associated with a CarD binding site (**Chapter 3**, **Table 1**). The genes within this set included both genes that were predicted to transcriptionally activated by CarD and genes that were predicted to be transcriptionally repressed by CarD. These data indicate that CarD binding *in vivo* is associated with transcriptional regulation but that binding alone does not determine the regulatory outcome.

CarD can directly activate or repress transcription depending on the kinetic context of a promoter

Our results thus far show that CarD is a bi-functional transcription regulator that is capable of either differentially activating or repressing transcription despite binding to DNA in a sequence-independent manner. These observations presented the intriguing question of how CarD could exert regulatory specificity without binding specificity. CarD's behavior as a transcriptional regulatory was reminiscent of transcription regulation by the bacterial stringent response factors DksA and guanine (penta-)tetraphosphate [(p)ppGpp], which bind directly to RNAP in *Escherichia coli* to activate transcription of amino acid biosynthesis genes and repress transcription of rRNA genes depending on the basal transcription initiation kinetics from these promoters (92, 94). Therefore, we hypothesized that CarD could discriminate mycobacterial promoters based on their basal kinetics and developed a model in collaboration with Eric Galburt's lab in which CarD activates transcription from promoters that form an unstable RP_o and represses transcription from promoters that form a basally stable RP_o (**Chapter 2**, **Figure 2**).

To test our model, I used the *Mtb rrnA*P3 promoter as a template to generate a set of promoter variants with a range of basal RP_o stabilities based on targeted DNA sequence mutations. Using our multi-round *in vitro* transcription assay and a three-nucleotide *in vitro* transcription assay designed to measure RP_o stability, I measured two characteristics of the AP3 promoter variants: 1) the basal RP_o stability of the promoter in the absence of CarD, and 2) the ratio of full-length RNA transcript produced in reactions with CarD versus without CarD. I show that sequence mutations that optimize DNA-RNAP interactions in RP_o (**Chapter 3**, **Figure 3**) or decrease the GC% of the promoter discriminator region (**Chapter 3**, **Figure 4**) increase basal RP_o stability. Furthermore, the AP3 promoter variants with the highest level of basal RP_o stability were transcriptionally repressed by CarD (**Chapter 3**, **Figure 3-4**), providing experimental evidence for our model of CarD's mechanism of promoter specificity. These data are also the first

demonstration of CarD directly repressing transcription *in vitro* using the mycobacterial transcription machinery. Relating the measurements in our *in vitro* transcription experiments reveals a strong positive correlation between basal RP_o stability and basal full-length transcript production and a negative correlation between basal RP_o stability and the ratio of full-length transcript production ±CarD (**Chapter 3**, **Figure 5**). Collectively, these data show that basal RP_o stability is a fundamental determinant of transcription strength and CarD regulation in mycobacterial promoters.

To investigate whether the mechanisms of promote specificity that we demonstrated *in vitro* were relevant to CarD's function *in vivo*, I analyzed the promoter sequences of genes that were differentially expressed in our *Mtb* and *M. smegmatis* RNA-seq datasets. My analysis shows that promoter characteristics indicative of high basal RP_o stability, such as the presence of an extended -10 motif and low discriminator region GC%, were associated with genes predicted to be repressed by CarD *in vivo* (**Chapter 3**, **Figure S4**). In addition, promoters lacking a T at position -12 in the template strand were significantly enriched in the set of genes predicted to be repressed by CarD (**Chapter 2**, **Figure 2**). On the surface, this association seems counterintuitive, because a non-consensus -10 element motif would be expected to have low RP_o stability and be activated by CarD. However, crystal structures of *Mtb* RP_o show that T_{-12} interacts with a conserved tryptophan residue in CarD to prevent transcription bubble collapse, so in theory these promoters that lack T_{-12} may be subject to a different regulatory mechanism.

Promoter DNA supercoiling influences its basal RP_o stability and CarD regulation

Our *in vitro* experiments using AP3 promoter variants demonstrate how promoter DNA sequence can influence basal RP_o stability and, in turn, CarD regulatory outcome. However, within a given cell, the DNA sequence at a given promoter is invariant but other environmental factors

such as DNA supercoiling are dynamic and can also influence RP_o stability (155). To test the effect of DNA supercoiling on CarD regulation, I generated a set of DNA templates with a fixed promoter sequence (WT *rrnA*P3) but encoded on either a negatively supercoiled plasmid, a nicked plasmid, a linearized plasmid, or a relaxed plasmid. Using *in vitro* transcription assays, I show that the negatively supercoiled promoter has a much greater basal RP_o stability and is transcriptionally repressed by CarD (**Chapter 3**, **Figure 6**). These data demonstrate that DNA sequence-independent factors such as DNA topology can also affect initiation kinetics and impact the outcome of CarD regulation.

The GlnR and ArgR regulons are indirectly repressed by CarD in *Mtb*

Using a statistical approach inspired by tests commonly used for gene enrichment analysis (169), I identified two *Mtb* transcription factor regulons that were significantly dysregulated in $CarD^{R47E}$ – GlnR and ArgR (**Chapter 4**, **Table 1**). My analysis suggests that CarD is represses the regulons of GlnR and ArgR, which activate transcription of the NirBD nitrite reductase operon and de novo arginine biosynthesis operon, respectively.

Open Questions

What is CarD's role in the mycobacterial stringent response?

While numerous kinetic studies and sequencing studies have elucidated CarD's regulatory mechanism and painted a broad-strokes image of CarD's role in mycobacterial growth and physiology, we still lack a detailed understanding of how transcription regulation by CarD translates into phenotypic changes in mycobacteria. Several studies also implicate CarD as an important mediator of various environmental stress responses including oxidative stress, nutrient starvation, genotoxic stress, and antibiotic treatment (63, 100–102). Based on the current evidence,

I believe that the most intriguing hypothesis of CarD's role in mycobacterial physiology is as an alternative stringent response regulator.

The stringent response is highly conserved nutrient starvation response in bacteria that canonically involves the synthesis of (p)ppGpp to down-regulate transcription of stable RNAs and up-regulate transcription of amino acid biosynthesis operons. However, details of the stringent response in *Mtb* are poorly understood (179). For example, *Mtb* does not encode a homolog of the stringent response mediator DksA and the *Mtb* RNAP lacks the (p)ppGpp binding sites characterized on the *E. coli* RNAP, suggesting that alternative stringent response regulators may be needed. Mechanistically, CarD shares many "anti-parallels" with DksA and (p)ppGpp (127). While DksA/(p)ppGpp destabilize an isomerization intermediate to activate transcription from promoters with stable RP_o and repress transcription from promoters with unstable RP_o, CarD does the opposite, stabilizing RP_o to activate transcription from promoters that form unstable RP_o and to repress transcription from promoters that form stable RP_o. In the canonical stringent response, (p)ppGpp is present at low levels during growth in nutrient replete conditions and then highly induced upon nutrient limitation; in contrast, CarD is expressed at high levels during growth in nutrient replete conditions and rapidly degraded upon nutrient limitation (102). DksA/(p)ppGpp repress transcription of rRNA promoters and activate transcription from amino acid biosynthesis operons; CarD activates transcription of rRNA promoters and indirectly represses transcription of the arginine biosynthesis pathway. Considering all of these facts together, CarD seems to function as a "reverse" stringent response regulator in *Mtb* compared to the canonical DksA/(p)ppGpp mediated pathway in *E. coli*.

Phenotypic studies of CarD mutants in nutrient starvation conditions yield conflicting results. In support the hypothesis that CarD functions as "reverse" stringent response regulator,

Mtb strains that cannot degrade CarD show reduced viability upon nutrient starvation (102). However, *Mtb* strains with CarD depleted also show reduced viability upon nutrient starvation (63), contrary to the expectations of this hypothesis. Collectively, these data suggest that proper regulation of intracellular CarD protein levels is critical for *Mtb* survival during nutrient starvation.

Thus far, all transcriptional studies of CarD in *Mtb* have been performed during log phase growth in nutrient replete conditions and in the context of the housekeeping σ^A holoenzyme. An interesting future direction that could help elucidate CarD's role in the *Mtb* starvation response is to perform transcriptional profiling of WT and CarD mutant *Mtb* cells entering starvation. The reliance on *Mtb* strains that encode point mutations in CarD also muddle the interpretation of *in vivo* transcriptomic data, because the individual mutations may have off-target effects on CarD's mechanism beyond RP_o stabilization. Future studies focus on using more direct methods to manipulate CarD activity, such as using CRISPR interference to knock down CarD expression or expressing the regulatory antisense RNA from an inducible promoter to titrate intracellular CarD levels.

What additional factors impact RP_o stability and CarD regulation?

Our experiments illustrate a fundamental connection between basal RP_o stability, transcription strength, and CarD regulation. The wide variance in transcriptional strength and basal RP_o stability at different bacterial promoters is largely explained by differences in DNA sequence (32, 34). However, our study also shows that two promoters with identical DNA sequences can have different levels of RP_o stability depending on additional factors such as DNA supercoiling. Unlike DNA sequence, the supercoiling status of a promoter can be dynamic over the lifetime of a bacterial cell, suggesting that the regulatory outcome at a given promoter by RNAP-binding transcription factors like CarD could also change in response to environmental stimuli. The

relationship between DNA supercoiling and CarD regulation is especially intriguing, because CarD transcript levels are up-regulated in response to DNA strand breaks in *Mtb* and CarD mutants are sensitized to DNA gyrase inhibitors such as ciprofloxacin (63, 100).

Extrapolating from this concept, any environmental factor that could alter RP_o stability could also change the direction of CarD regulation at a given promoter. For example, intracellular nucleotide triphosphate (NTP) concentrations regulate the transcription of rRNA promoters in *E. coli*, because the binding of the initiating NTP to the RNAP active site stabilizes RP_o (162). Thus, CarD-regulated promoters in *Mtb* could in theory sense and respond to environmental cues like NTP availability without changes in CarD expression or activity. In addition to DNA supercoiling and NTP concentration, other factors such as temperature, salt conditions, and local binding of nucleoid-associated proteins could all dynamically alter RP_o stability and CarD regulation. Future *in vitro* transcription studies can explore this concept by interrogating CarD's regulatory activity in different environmental contexts.

How has CarD co-evolved with the genome and transcriptional machinery of other bacterial species?

CarD homologs are widely distributed in many bacterial phyla, but its function and essentiality vary across species. For example, in *Mtb* CarD is essential and regulates a broad range of genes but in *Borrelia burgdorferi* CarD is non-essential and seems to specifically regulate a set of genes involved in bacterial growth at low temperatures (109). Furthermore, many bacteria like *E. coli* do not even encode for a homolog of CarD, suggesting that the essential functions that it plays in *Mtb* are either not necessary in *E. coli* or compensated for by a different mechanism. Currently, it remains unclear why CarD is essential in some bacteria but not in others.

One potential hypothesis is that the essentiality of CarD has evolved with the GC-richness of the genome. Our data shows that CarD is required by the *Mtb* RNAP to activate transcription from GC-rich promoters that form unstable RP_o. Therefore, one might expect that CarD would be essential in bacteria that have genomes with high GC-content. While this prediction holds true for many bacteria including *Mtb* (essential; GC-content 65.6%), *Myxococcus xanthus* (essential; GC-content 68.9%), *B. burgdorferi* (non-essential; GC-content 28.6%), and *Bacillus cereus* (non-essential; GC-content 35.4%), certain bacteria such as *Caulobacter crescentus* (non-essential; GC-content 67.2%) break this trend. However, this discrepancy does not disprove this hypothesis, because the mean GC-content of a genome does not necessarily reflect the GC-content of relevant promoter regions that may be regulated by CarD. Furthermore, as our studies of the *Mtb* RNAP have highlighted, the transcription paradigms developed from studies of one species may not be applicable to the transcription machinery in other bacteria. For example, *C. crescentus* utilizes the non-canonical transcription initiation factor GcrA to regulate its unique cell cycle progression (180). In addition, CarD may function differently in other bacteria. In *M. smegmatis*, CarD bound to a majority of the promoters that were differentially expressed in the CarD mutant strains, but in *C. crescentus* a similar analysis showed that a vast majority of the promoters differentially expressed in the Δ*cdnL* (CarD homolog in *C. crescentus*) strain are not bound by CdnL (107). The question 'why is CarD is essential in some bacteria but not others?' is difficult to address experimentally, and the answer may naturally appear as future studies characterize CarD homologs and unique transcription initiation paradigms in more bacterial species.

References

1. Zimmerman, M. R. (1979) Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bull N Y Acad Med*. **55**, 604–608
2. Barberis, I., Bragazzi, N. L., Galluzzo, L., and Martini, M. (2017) The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus. *J. Prev. Med. Hyg*. **58**, E9–E12
3. Koch, R. (1882) Die Aetiologie der Tuberkulose. *Berliner Klin. Wochenschrift*
4. World Health Organization (2021) *Global tuberculosis report 2021*, Geneva
5. Gagneux, S. (2018) Ecology and evolution of Mycobacterium tuberculosis. *Nat. Rev. Microbiol*. **16**, 202–213
6. Boritsch, E. C., and Brosch, R. (2017) Evolution of Mycobacterium tuberculosis: New insights into pathogenicity and drug resistance. *Tuberc. Tuber. Bacillus Second Ed*. 10.1128/9781555819569.ch22
7. Veyrier, F., Pletzer, D., Turenne, C., and Behr, M. A. (2009) Phylogenetic detection of horizontal gene transfer during the step-wise genesis of Mycobacterium tuberculosis. *BMC Evol. Biol*. **9**, 1–14
8. Sala, A., Bordes, P., and Genevaux, P. (2014) Multiple toxin-antitoxin systems in Mycobacterium tuberculosis. *Toxins (Basel)*. **6**, 1002–1020
9. Dunlap, M. D., Howard, N., Das, S., Scott, N., Ahmed, M., Prince, O., Rangel-Moreno, J., Rosa, B. A., Martin, J., Kaushal, D., Kaplan, G., Mitreva, M., Kim, K.-W., Randolph, G. J., and Khader, S. A. (2018) A novel role for C–C motif chemokine receptor 2 during infection with hypervirulent Mycobacterium tuberculosis. *Mucosal Immunol*. **11**, 1727–1742

10. Ryndak, M. B., and Laal, S. (2019) Mycobacterium tuberculosis Primary Infection and Dissemination: A Critical Role for Alveolar Epithelial Cells. *Front. Cell. Infect. Microbiol.* 10.3389/fcimb.2019.00299

11. Cohen, S. B., Gern, B. H., Delahaye, J. L., Adams, K. N., Plumlee, C. R., Winkler, J. K., Sherman, D. R., Gerner, M. Y., and Urdahl, K. B. (2018) Alveolar Macrophages Provide an Early Mycobacterium tuberculosis Niche and Initiate Dissemination. *Cell Host Microbe*. **24**, 439-446.e4

12. Chandra, P., Grigsby, S. J., and Philips, J. A. (2022) Immune evasion and provocation by Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **20**, 750–766

13. Stallings, C. L., and Glickman, M. S. (2010) Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. *Microbes Infect.* **12**, 1091–1101

14. Kinsella, R. L., Zhu, D. X., Harrison, G. A., Mayer Bridwell, A. E., Prusa, J., Chavez, S. M., and Stallings, C. L. (2021) Perspectives and Advances in the Understanding of Tuberculosis. *Annu. Rev. Pathol. Mech. Dis.* **16**, 377–408

15. Vandal, O. H., Pierini, L. M., Schnappinger, D., Nathan, C. F., and Ehrt, S. (2008) A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. *Nat. Med.* **14**, 849–854

16. Miller, J. L., Velmurugan, K., Cowan, M. J., and Briken, V. (2010) The Type I NADH Dehydrogenase of Mycobacterium tuberculosis Counters Phagosomal NOX2 Activity to Inhibit TNF-α-Mediated Host Cell Apoptosis. *PLOS Pathog.* **6**, e1000864

17. Pagan, A. J., and Ramakrishnan, L. (2014) Immunity and Immunopathology in the Tuberculous Granuloma. *Cold Spring Harb. Perspect. Med.* 10.1101/cshperspect.a018499

18. Marakalala, M. J., Raju, R. M., Sharma, K., Zhang, Y. J., Eugenin, E. A., Prideaux, B.,

Daudelin, I. B., Chen, P., Booty, M. G., Kim, J. H., Eum, S. Y., Via, L. E., Behar, S. M., Iii, C. E. B., Mann, M., Dartois, V., and Rubin, E. J. (2016) Inflammatory signaling in human tuberculosis granulomas is spatially organized. 10.1038/nm.4073

19. Martin, C. J., Cadena, A. M., Leung, V. W., Lin, P. L., Maiello, P., Hicks, N., Chase, M. R., Flynn, J. A. L., and Fortune, S. M. (2017) Digitally barcoding Mycobacterium tuberculosis reveals In vivo infection dynamics in the macaque model of tuberculosis. *MBio*. 10.1128/mBio.00312-17

20. Lin, P. L., Maiello, P., Gideon, H. P., Coleman, M. T., Cadena, M., Rodgers, M. A., Gregg, R., Malley, M. O., Tomko, J., Fillmore, D., Frye, L. J., Rutledge, T., Difazio, R. M., Janssen, C., Klein, E., Andersen, P. L., Fortune, S. M., and Flynn, L. (2016) PET CT Identifies Reactivation Risk in Cynomolgus Macaques with Latent M . tuberculosis. 10.1371/journal.ppat.1005739

21. Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V., and Andersen, Å. B. (2002) Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. *J. Infect. Dis.* **185**, 401–404

22. Behr, M. A., Edelstein, P. H., and Ramakrishnan, L. (2018) Revisiting the timetable of tuberculosis. *BMJ*. **362**, 1–10

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

24. Murakami, K. S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules*. **5**, 848–864

25. Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F.,

Glickman, M. S., Landick, R., Darst, S. A., Stallings, C. L., and Campbell, E. A. (2013) Structure and function of CarD, an essential mycobacterial transcription factor. *Proc. Natl. Acad. Sci.* **110**, 12619–12624

26. Boyaci, H., Chen, J., Jansen, R., Darst, S. A., and Campbell, E. A. (2019) Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature*. **565**, 382–385

27. Hubin, E. A., Lilic, M., Darst, S. A., and Campbell, E. A. (2017) Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nat. Commun.* **8**, 1–12

28. Hubin, E. A., Fay, A., Xu, C., Bean, J. M., Saecker, R. M., Glickman, M. S., Darst, S. A., and Campbell, E. A. (2017) Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife*. **6**, 1–40

29. Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y. W., Li, S., Feng, Y., Zhang, Y., Mandal, S., Jiang, Y., Liu, S., Gigliotti, M., Talaue, M., Connell, N., Das, K., Arnold, E., and Ebright, R. H. (2017) Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Mol. Cell*. **66**, 169-179.e8

30. Mao, C., Zhu, Y., Lu, P., Feng, L., Chen, S., and Hu, Y. (2018) Association of ω with the Cterminal region of the β' subunit is essential for assembly of RNA polymerase in Mycobacterium tuberculosis. *J. Bacteriol*. 10.1128/JB.00159-18

31. Nickels, B. E., and Hochschild, A. (2004) Regulation of RNA polymerase through the secondary channel. *Cell*. **118**, 281–284

32. Lee, J., and Borukhov, S. (2016) Bacterial RNA polymerase-DNA interaction-The driving force of gene expression and the target for drug action. *Front. Mol. Biosci.*

10.3389/fmolb.2016.00073

33. Martin, C. T., and Theis, K. (2014) Closed for business: Exit-channel coupling to active site conformation in bacterial RNA polymerase. *Nat. Struct. Mol. Biol.* **21**, 741–742

34. Jensen, D., and Galburt, E. A. (2021) The Context-Dependent Influence of Promoter Sequence Motifs on Transcription Initiation Kinetics and Regulation. *J. Bacteriol.* **203**, 1–18

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

36. Dombroski, A. J., Walter, W. A., and Gross, C. A. (1993) Amino-terminal amino acids modulate σ-factor DNA-binding activity. *Genes Dev.* **7**, 2446–2455

37. Murakami, K. S. (2013) X-ray crystal structure of escherichia coli RNA polymerase σ70 holoenzyme. *J. Biol. Chem.* **288**, 9126–9134

38. Ruff, E. F., Thomas Record, M., and Artsimovitch, I. (2015) Initial events in bacterial transcription initiation. *Biomolecules*. **5**, 1035–1062

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

40. Chen, J., Boyaci, H., and Campbell, E. A. (2021) Diverse and unified mechanisms of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **19**, 95–109

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

102, 4706–4711

42. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993) A Third Recognition Element in Bacterial Promoters: DNA Binding by the α Subunit of RNA Polymerase. *Science (80-.).* **262**, 1407–1413

43. Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S., and Ebright, R. H. (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science (80-.).* **337**, 591–595

44. Feklistov, A., Bae, B., Hauver, J., Lass-Napiorkowska, A., Kalesse, M., Glaus, F., Altmann, K.-H., Heyduk, T., Landick, R., and Darst, S. A. (2017) PROTEIN DYNAMICS RNA polymerase motions during promoter melting. *Science (80-.).* **356**, 863–866

45. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S. A. (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife*. **4**, 1–23

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

47. Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. (2006) Abortive Initiation and Productive Initiation by RNA Polyermase Involve DNA Scrunching. *Science (80-.).* **314**, 1139–1143

48. Li, L., Molodtsov, V., Lin, W., Ebright, R. H., and Zhang, Y. (2020) RNA extension drives a stepwise displacement of an initiation-factor structural module in initial

transcription. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5801–5809

49. Henderson, K. L., Felth, L. C., Molzahn, C. M., Shkel, I., Wang, S., Chhabra, M., Ruff, E. F., Bieter, L., Kraft, J. E., and Record, M. T. (2017) Mechanism of transcription initiation and promoter escape by *E . coli* RNA polymerase. *Proc. Natl. Acad. Sci.* **114**, E3032–E3040

50. Susa, M., Sen, R., and Shimamoto, N. (2002) Generality of the branched pathway in transcription initiation by Escherichia coli RNA polymerase. *J. Biol. Chem.* **277**, 15407–15412

51. Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., Winkelman, J. T., Maloney, M. F., Chait, B. T., Ross, W., Gourse, R. L., Campbell, E. A., and Darst, S. A. (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol. Cell.* **78**, 275-288.e6

52. Mazumder, A., and Kapanidis, A. N. (2019) Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *J. Mol. Biol.* **431**, 3947–3959

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

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

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

56. Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., and Ebright, R. H.

(2012) Structural Basis of Transcription Initiation. *Science (80-.).* **338**, 1076–1080

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

58. Myers, K. S., Noguera, D. R., and Donohue, T. J. (2021) Promoter Architecture Differences among Alphaproteobacteria and Other Bacterial Taxa . *mSystems*. 10.1128/msystems.00526-21

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

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

61. Zhu, Y., Mao, C., Ge, X., Wang, Z., and Lu, P. (2017) Characterization of a Minimal Type of Position in Mycobacteria. **199**, 1–12

62. Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.-P., and Brodolin, K. (2012) Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the σ A-containing RNA polymerase holoenzyme. *Nucleic Acids Res*. **40**, 6547–57

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

64. Rammohan, J., Ruiz Manzano, A., Garner, A. L., Prusa, J., Stallings, C. L., and Galburt, E. A. (2016) Cooperative stabilization of Mycobacterium tuberculosis rrnA P3 promoter open complexes by RbpA and CarD. *Nucleic Acids Res*. **44**, gkw577

65. Rodrigue, S., Provvedi, R., Jacques, P. É., Gaudreau, L., and Manganelli, R. (2006) The ?? factors of Mycobacterium tuberculosis. *FEMS Microbiol. Rev.* **30**, 926–941

66. Manganelli, R. (2015) Sigma factors: Key molecules in Mycobacterium tuberculosis physiology and virulence. *Mol. Genet. Mycobact.* 10.1128/9781555818845.ch7

67. Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L., Smith, I., and Proveddi, R. (2004) Sigma factors and global gene regulation in Mycobacterium tuberculosis. *J. Bacteriol.* **186**, 895–902

68. Flentie, K., Garner, A. L., and Stallings, C. L. (2016) Mycobacterium tuberculosis Transcription Machinery: Ready To Respond to Host Attacks. *J. Bacteriol.* **198**, 1360–1373

69. Vishwakarma, R. K., and Brodolin, K. (2020) The σ Subunit-Remodeling Factors: An Emerging Paradigms of Transcription Regulation. *Front. Microbiol.* **11**, 1–9

70. Bortoluzzi, A., Muskett, F. W., Waters, L. C., Addis, P. W., Rieck, B., Munder, T., Schleier, S., Forti, F., Ghisotti, D., Carr, M. D., and O'Hare, H. M. (2013) Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. *J. Biol. Chem.* **288**, 14438–14450

71. Perumal, A. S., Vishwakarma, R. K., Hu, Y., Morichaud, Z., and Brodolin, K. (2018) RbpA relaxes promoter selectivity of M. tuberculosis RNA polymerase. *Nucleic Acids Res.* **46**, 10106–10118

72. Vishwakarma, R. K., Cao, A., Morichaud, Z., Perumal, A. S., Margeat, E., and Brodolin, K. (2018) Single-molecule analysis reveals the mechanism of transcription activation in M. tuberculosis. *Sci. Adv.* 10.1126/sciadv.aao5498

73. Hubin, E. A., Tabib-Salazar, A., Humphrey, L. J., Flack, J. E., Olinares, P. D. B., Darst, S.

A., Campbell, E. A., and Paget, M. S. (2015) Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc. Natl. Acad. Sci.* **112**, 7171 LP – 7176

74. Prusa, J., Jensen, D., Santiago-Collazo, G., Pope, S. S., Garner, A. L., Miller, J. J., Ruiz Manzano, A., Galburt, E. A., and Stallings, C. L. (2018) Domains within RbpA serve specific functional roles that regulate the expression of distinct mycobacterial gene subsets. *J. Bacteriol.*

75. Jensen, D., Manzano, A. R., Rammohan, J., Stallings, C. L., and Galburt, E. A. (2019) CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. *Nucleic Acids Res.* **47**, 6685–6698

76. Prusa, J., Zhu, D. X., Flynn, A. J., Jensen, D., Manzano, A. R., Galburt, E. A., and Stallings, C. L. (2022) Molecular dissection of RbpA-mediated regulation of fidaxomicin sensitivity in mycobacteria. *J. Biol. Chem.* **298**, 101752

77. Boyaci, H., Chen, J., Lilic, M., Palka, M., Mooney, R. A., Landick, R., Darst, S. A., and Campbell, E. A. (2018) Fidaxomicin jams mycobacterium tuberculosis RNA polymerase motions needed for initiation via RBPA contacts. *Elife.* **7**, 1–19

78. Browning, D. F., Butala, M., and Busby, S. J. W. (2019) Bacterial Transcription Factors: Regulation by Pick "N" Mix. *J. Mol. Biol.* **431**, 4067–4077

79. Browning, D. F., and Busby, S. J. W. (2016) Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650

80. Minch, K. J., Rustad, T. R., Peterson, E. J. R., Winkler, J., Reiss, D. J., Ma, S., Hickey, M., Brabant, W., Morrison, B., Turkarslan, S., Mawhinney, C., Galagan, J. E., Price, N. D., Baliga, N. S., and Sherman, D. R. (2015) The DNA-binding network of

Mycobacterium tuberculosis. *Nat. Commun.* **6**, 5829

81. Galagan, J. E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., Dolganov, G., Glotova, I., Abeel, T., Mahwinney, C., Kennedy, A. D., Allard, R., Brabant, W., Krueger, A., Jaini, S., Honda, B., Yu, W. H., Hickey, M. J., Zucker, J., Garay, C., Weiner, B., Sisk, P., Stolte, C., Winkler, J. K., Van De Peer, Y., Iazzetti, P., Camacho, D., Dreyfuss, J., Liu, Y., Dorhoi, A., Mollenkopf, H. J., Drogaris, P., Lamontagne, J., Zhou, Y., Piquenot, J., Park, S. T., Raman, S., Kaufmann, S. H. E., Mohney, R. P., Chelsky, D., Branch Moody, D., Sherman, D. R., and Schoolnik, G. K. (2013) The Mycobacterium tuberculosis regulatory network and hypoxia. *Nature*. **499**, 178–183

82. Browning, D. F., Grainger, D. C., and Busby, S. J. W. (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* **13**, 773–780

83. Bush, M. J. (2018) The actinobacterial WhiB-like (Wbl) family of transcription factors. *Mol. Microbiol.* **110**, 663–676

84. Lilic, M., Darst, S. A., and Campbell, E. A. (2021) Structural basis of transcriptional activation by the Mycobacterium tuberculosis intrinsic antibiotic-resistance transcription factor WhiB7. *Mol. Cell.* **81**, 2875-2886.e5

85. Parish, T. (2015) Two-component regulatory systems of mycobacteria. *Mol. Genet. Mycobact.* 10.1128/9781555818845.ch10

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

87. Cavanagh, A. T., and Wassarman, K. M. (2014) 6S RNA, a global regulator of transcription in escherichia coli, bacillus subtilis, and beyond. *Annu. Rev. Microbiol.* **68**, 45–60

88. Irving, S. E., Choudhury, N. R., and Corrigan, R. M. (2021) The stringent response and physiological roles of (pp)pGpp in bacteria. *Nat. Rev. Microbiol.* **19**, 256–271

89. Gourse, R. L., Chen, A. Y., Gopalkrishnan, S., Sanchez-Vazquez, P., Myers, A., and Ross, W. (2018) Transcriptional Responses to ppGpp and DksA. *Annu. Rev. Microbiol.* **72**, 163–184

90. Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T., and Conway, T. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. *Mol. Microbiol.* **68**, 1128–1148

91. Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S. E., and Murakami, K. S. (2018) Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol. Cell.* **69**, 828-839.e5

92. Paul, B. J., Berkmen, M. B., and Gourse, R. L. (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc. Natl. Acad. Sci.* **102**, 7823–7828

93. Paul, B. J., Barker, M. M., Ross, W., Schneider, D. A., Webb, C., Foster, J. W., and Gourse, R. L. (2004) DksA: A critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell.* **118**, 311–322

94. Rutherford, S. T., Villers, C. L., Lee, J., Ross, W., and Gourse, R. L. (2009) Allosteric control of Escherichia coli rRNA promoter complexes by DksA. 10.1101/gad.1745409.Because

95. Sanchez-Vazquez, P., Dewey, C. N., Kitten, N., Ross, W., and Gourse, R. L. (2019) Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci.* **116**, 8310 LP – 8319

96. Bae, B., Chen, J., Davis, E., Leon, K., Darst, S. A., and Campbell, E. A. (2015) CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife*. **4**, 1–19

97. Weiss, L. A., Harrison, P. G., Nickels, B. E., Glickman, M. S., Campbell, E. A., Darst, S. A., and Stallings, C. L. (2012) Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. *J. Bacteriol.* **194**, 5621–5631

98. Gulten, G., and Sacchettini, J. C. (2013) Structure of the Mtb CarD/RNAP β-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure*. **21**, 1859–1869

99. Westblade, L. F., Campbell, E. A., Pukhrambam, C., Padovan, J. C., Nickels, B. E., Lamour, V., and Darst, S. A. (2010) Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res*. **38**, 8357–8369

100. Garner, A. L., Weiss, L. A., Manzano, A. R., Galburt, E. A., and Stallings, C. L. (2014) CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. *Mol. Microbiol*. **93**, 682–697

101. Garner, A. L., Rammohan, J., Huynh, J. P., Onder, L. M., Chen, J., Bae, B., Jensen, D., Weiss, L. A., Manzano, A. R., Darst, S. A., Campbell, E. A., Nickels, B. E., Galburt, E. A., and Stallings, C. L. (2017) Effects of increasing the affinity of CarD for RNA polymerase on Mycobacterium tuberculosis growth, rRNA transcription, and virulence. *J.*

Bacteriol. 10.1128/JB.00698-16

102. Li, X., Chen, F., Liu, X., Xiao, J., Andongma, B. T., Tang, Q., and Cao, X. (2022) Clp protease and antisense RNA jointly regulate the global regulator CarD to mediate mycobacterial starvation response

103. Landick, R., Krek, A., Glickman, M. S., Socci, N. D., and Stallings, C. L. (2014) Genome-Wide Mapping of the Distribution of CarD, RNAP σ(A), and RNAP β on the Mycobacterium smegmatis Chromosome using Chromatin Immunoprecipitation Sequencing. *Genomics data*. **2**, 110–113

104. García-Moreno, D., Abellón-Ruiz, J., García-Heras, F., Murillo, F. J., Padmanabhan, S., and Elías-Arnanz, M. (2010) CdnL, a member of the large CarD-like family of bacterial proteins, is vital for Myxococcus xanthus and differs functionally from the global transcriptional regulator CarD. *Nucleic Acids Res*. **38**, 4586–4598

105. Gallego-García, A., Mirassou, Y., García-Moreno, D., Elías-Arnanz, M., Jiménez, M. A., and Padmanabhan, S. (2014) Structural insights into RNA polymerase recognition and essential function of Myxococcus xanthus CdnL. *PLoS One*. 10.1371/journal.pone.0108946

106. Gallego-García, A., Iniesta, A. A., González, D., Collier, J., Padmanabhan, S., and Elías-Arnanz, M. (2017) Caulobacter crescentus CdnL is a non-essential RNA polymerase-binding protein whose depletion impairs normal growth and rRNA transcription. *Sci. Rep*. **7**, 43240

107. Woldemeskel, S. A., Daitch, A. K., Alvarez, L., Panis, G., Zeinert, R., Gonzalez, D., Smith, E., Collier, J., Chien, P., Cava, F., Viollier, P. H., and Goley, E. D. (2020) The conserved transcriptional regulator CdnL is required for metabolic homeostasis and

morphogenesis in Caulobacter. *PLoS Genet*. **16**, 1–25

108. Warda, A. K., Tempelaars, M. H., Boekhorst, J., Abee, T., and Nierop Groot, M. N. (2016) Identification of CdnL, a Putative Transcriptional Regulator Involved in Repair and Outgrowth of Heat-Damaged Bacillus cereus Spores. *PLoS One*. **11**, e0148670

109. Chen, T., Xiang, X., Xu, H., Zhang, X., Zhou, B., Yang, Y., Lou, Y., and Yang, X. F. (2018) LtpA, a CdnL-type CarD regulator, is important for the enzootic cycle of the Lyme disease pathogen. *Emerg. Microbes Infect*. **7**, 1–9

110. Henry, K. K., Ross, W., Myers, K. S., Lemmer, K. C., Vera, J. M., Landick, R., Donohue, T. J., and Gourse, R. L. (2020) A majority of Rhodobacter sphaeroides promoters lack a crucial RNA polymerase recognition feature, enabling coordinated transcription activation. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 29658–29668

111. Zhu, D. X., Garner, A. L., Galburt, E. A., and Stallings, C. L. (2019) CarD contributes to diverse gene expression outcomes throughout the genome of Mycobacterium tuberculosis. **116**, 13573–13581

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

113. Mekler, V., Minakhin, L., Kuznedelov, K., Mukhamedyarov, D., and Severinov, K. (2012) RNA polymerase-promoter interactions determining different stability of the Escherichia coli and Thermus aquaticus transcription initiation complexes. *Nucleic Acids Res*. **40**, 11352–11362

114. Gomez, M., and Smith, I. (2000) *Determinants of Mycobacterial Gene Expression*

115. WHO (2017) *Global Tuberculosis Report 2017*, WHO/HTM/TB/2017.23

116. Pan, L., da Silva, D., Pagliai, F. A., Harrison, N. A., Gonzalez, C. F., and Lorca, G. L.

(2019) The Ferredoxin-Like Protein FerR Regulates PrbP Activity in Liberibacter asiaticus. *Appl. Environ. Microbiol.* **85**, e02605-18

117. Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527

118. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21

119. Soneson, C., Love, M. I., and Robinson, M. D. (2018) Differential analyses for RNA-seq : transcript-level estimates improve gene-level inferences [version 1 ; referees : 2 approved] Referee Status : 10.12688/f1000research.7563.1

120. Zhu, A., Ibrahim, J. G., and Love, M. I. (2018) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *bioRxiv*

121. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210

122. Hsu, L. M. B. T.-M. in E. (1996) [4] Quantitative parameters for promoter clearance. in *RNA Polymerase and Associated Factors Part A*, pp. 59–71, Academic Press, **273**, 59–71

123. Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001) Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro11Edited by R. Ebright. *J. Mol. Biol.* **305**, 673–688

124. Lew, J. M., Kapopoulou, A., Jones, L. M., and Cole, S. T. (2011) TubercuList – 10 years after. *Tuberculosis.* **91**, 1–7

125. Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J. K. (2016) The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol. Cell. Biol.* **36**, 662–667

126. Doniselli, N., Rodriguez-Aliaga, P., Amidani, D., Bardales, J. A., Bustamante, C., Guerra, D. G., and Rivetti, C. (2015) New insights into the regulatory mechanisms of ppGpp and DksA on Escherichia coli RNA polymerase–promoter complex. *Nucleic Acids Res*. **43**, 5249–5262

127. Galburt, E. A. (2018) The calculation of transcript flux ratios reveals single regulatory mechanisms capable of activation and repression. *Proc. Natl. Acad. Sci*. **115**, E11604 LP-E11613

128. Hsu, L. M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta*. **1577**, 191–207

129. Siwo, G., Rider, A., Tan, A., Pinapati, R., Emrich, S., Chawla, N., and Ferdig, M. (2016) Prediction of fine-tuned promoter activity from DNA sequence. *F1000Research*. **5**, 158

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

131. Djordjevic, M., and Bundschuh, R. (2008) Formation of the open complex by bacterial RNA polymerase--a quantitative model. *Biophys. J*. **94**, 4233–4248

132. Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., and Young, D. B. (2013) Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in Mycobacterium tuberculosis. *Cell Rep*. **5**, 1121–1131

133. Bashyam, M. D., and Tyagi, A. K. (1998) Identification and Analysis of "Extended -10" Promoters from Mycobacteria. *J. Bacteriol*. **180**, 2568–73

134. Agarwal, N., and Tyagi, A. K. (2003) Role of 5′-TGN-3′ motif in the interaction of mycobacterial RNA polymerase with a promoter of 'extended —10' class. *FEMS Microbiol. Lett*. **225**, 75–83

135. Agarwal, N., and Tyagi, A. K. (2006) Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity. *Nucleic Acids Res*. **34**, 4245–57

136. Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., Gawande, R., Ahmad, R., Sarracino, D. A., Ioerger, T. R., Fortune, S. M., Derbyshire, K. M., Wade, J. T., and Gray, T. A. (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLOS Genet*. **11**, e1005641

137. Heyduk, E., and Heyduk, T. (2018) DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Res*. **46**, 4469–4486

138. Bashyam, M. D., Kaushal, D., Dasgupta, S. K., and Tyagi, A. K. (1996) A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J. Bacteriol*. **178**, 4847–53

139. Sanchez-Vazquez, P., Dewey, C. N., Kitten, N., Ross, W., and Gourse, R. L. (2019) Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A*. **116**, 8310–8319

140. Henry, K. K., Ross, W., and Gourse, R. L. (2021) Rhodobacter sphaeroides card negatively regulates its own promoter. *J. Bacteriol*. 10.1128/JB.00210-21

141. Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120

142. Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol*. **37**, 907–915

143. Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M.,

Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C. G., Genez, T., Martinez, J. G., Guijarro-Clarke, C., Gymer, A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Marugán, J. C., Mohanan, S., Mushtaq, A., Naven, M., Ogeh, D. N., Parker, A., Parton, A., Perry, M., Pilizota, I., Prosovetskaia, I., Sakthivel, M. P., Salam, A. I. A., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Perez-Silva, J. G., Stark, W., Steed, E., Sutinen, K., Sukumaran, R., Sumathipala, D., Suner, M. M., Szpak, M., Thormann, A., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, T. A., Walts, B., Willhoft, N., Winterbottom, A., Wass, E., Chakiachvili, M., Flint, B., Frankish, A., Giorgetti, S., Haggerty, L., Hunt, S. E., Iisley, G. R., Loveland, J. E., Martin, F. J., Moore, B., Mudge, J. M., Muffato, M., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, S. J., Dyer, S., Harrison, P. W., Howe, K. L., Yates, A. D., Zerbino, D. R., and Flicek, P. (2022) Ensembl 2022. *Nucleic Acids Res*. **50**, D988–D995

144. Liao, Y., Smyth, G. K., and Shi, W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930

145. Banerjee, R., Rudra, P., Prajapati, R. K., Sengupta, S., and Mukhopadhyay, J. (2014) Optimization of recombinant Mycobacterium tuberculosis RNA polymerase expression and purification. *Tuberculosis*. **94**, 397–404

146. Martini, M. C., Zhou, Y., Sun, H., and Shell, S. S. (2019) Defining the transcriptional and post-transcriptional landscapes of mycobacterium smegmatisin aerobic growth and hypoxia. *Front. Microbiol*. **10**, 1–17

147. Keilty, S., and Rosenberg, M. (1987) Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.* **262**, 6389–6395

148. Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimašauskas, S., Nikiforov, V., Heyduk, T., Severinov, K., and Kulbachinskiy, A. (2006) A Basal Promoter Element Recognized by Free RNA Polymerase σ Subunit Determines Promoter Recognition by RNA Polymerase Holoenzyme. *Mol. Cell.* **23**, 97–107

149. Barinova, N., Kuznedelov, K., Severinov, K., and Kulbachinskiy, A. (2008) Structural modules of RNA polymerase required for transcription from promoters containing downstream basal promoter element GGGA. *J. Biol. Chem.* **283**, 22482–22489

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

151. Yakovchuk, P., Protozanova, E., and Frank-Kamenetskii, M. D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574

152. Wang, A. H. J., Hakoshima, T., van der Marel, G., van Boom, J. H., and Rich, A. (1984) AT base pairs are less stable than GC base pairs in Z-DNA: The crystal structure of d(m5CGTAm5CG). *Cell.* **37**, 321–331

153. Pemberton, I. K., Muskhelishvili, G., Travers, A. A., Buckle, M., and Physicochimie, Â. De (2000) The G+C-rich Discriminator Region of the tyrT Promoter Antagonises the Formation of Stable Preinitiation Complexes. 10.1006/jmbi.2000.3780

154. Gummesson, B., Lovmar, M., and Nyström, T. (2013) A proximal promoter element required for positive transcriptional control by guanosine tetraphosphate and DKSA

protein during the stringent response. *J. Biol. Chem.* **288**, 21055–21064

155. Dorman, C. J. (2019) DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. cell Biol.* **20**, 26

156. Revyakin, A., Ebright, R. H., and Strick, T. R. (2004) Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4776–4780

157. Saecker, R. M., Chen, J., Chiu, C. E., Malone, B., Sotiris, J., Ebrahim, M., Yen, L. Y., Eng, E. T., and Darst, S. A. (2021) Structural origins of Escherichia coli RNA polymerase open promoter complex stability. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–9

158. Petushkov, I., Pupov, D., Bass, I., and Kulbachinskiy, A. (2015) Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Res.* **43**, 5798–5809

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

160. Dulin, D., Bauer, D. L. V., Malinen, A. M., Bakermans, J. J. W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., Kulbachinskiy, A., and Kapanidis, A. N. (2018) Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nat. Commun.* **9**, 1–12

161. Dorman, C. J., and Dorman, M. J. (2016) DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys. Rev.* **8**, 89–100

162. Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L., and Gourse, R. L. (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science (80-.).* **278**, 2092–2097

163. Plaskon, D. M., Henderson, K. L., Felth, L. C., Molzahn, C. M., Evensen, C., Dyke, S., Shkel, I. A., and Thomas Record, M. (2021) Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–8

164. Chen, J., Gopalkrishnan, S., Chiu, C., Chen, A. Y., Campbell, E. A., Gourse, R. L., Ross, W., and Darst, S. A. (2019) E. Coli trar allosterically regulates transcription initiation by altering RNA polymerase conformation. *Elife*. **8**, 1–29

165. Gopalkrishnan, S., Ross, W., Akbari, M. S., Li, X., Haycocks, J. R. J., Grainger, D. C., Court, D. L., and Gourse, R. L. (2022) Homologs of the Escherichia coli F Element Protein TraR, Including Phage Lambda Orf73, Directly Reprogram Host Transcription. *MBio*. **13**, e0095222

166. Lemke, J. J., Sanchez-Vazquez, P., Burgos, H. L., Hedberg, G., Ross, W., and Gourse, R. L. (2011) Direct regulation of Escherichia coli ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc. Natl. Acad. Sci.* **108**, 5712

167. Lunge, A., Gupta, R., Choudhary, E., and Agarwal, N. (2020) The unfoldase ClpC1 of mycobacterium tuberculosis regulates the expression of a distinct subset of proteins having intrinsically disordered termini. *J. Biol. Chem.* **295**, 9455–9473

168. Rustad, T. R., Minch, K. J., Ma, S., Winkler, J. K., Hobbs, S., Hickey, M., Brabant, W., Turkarslan, S., Price, N. D., Baliga, N. S., and Sherman, D. R. (2014) Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biol.* **15**, 502

169. Matern, W. M., Rifat, D., Bader, J. S., and Karakousis, P. C. (2018) Gene enrichment analysis reveals major regulators of Mycobacterium tuberculosis gene expression in two

models of antibiotic tolerance. *Front. Microbiol.* **9**, 1–10

170. Turkarslan, S., Peterson, E. J. R., Rustad, T. R., Minch, K. J., Reiss, D. J., Morrison, R., Ma, S., Price, N. D., Sherman, D. R., and Baliga, N. S. (2015) A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis. *Sci. Data.* **2**, 1–10

171. Fink, D., Weißschuh, N., Reuther, J., Wohlleben, W., and Engels, A. (2002) Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in Streptomyces coelicolor A3(2). *Mol. Microbiol.* **46**, 331–347

172. Jenkins, V. A., Barton, G. R., Robertson, B. D., and Williams, K. J. (2013) Genome wide analysis of the complete GlnR nitrogen-response regulon in Mycobacterium smegmatis. *BMC Genomics.* **14**, 1–15

173. Amon, J., Bräu, T., Grimrath, A., Hänßler, E., Hasselt, K., Höller, M., Jeßberger, N., Ott, L., Szököl, J., Titgemeyer, F., and Burkovski, A. (2008) Nitrogen control in mycobacterium smegmatis: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-Type regulator GlnR. *J. Bacteriol.* **190**, 7108–7116

174. Malm, S., Tiffert, Y., Micklinghoff, J., Schultze, S., Joost, I., Weber, I., Horst, S., Ackermann, B., Schmidt, M., Wohlleben, W., Ehlers, S., Geffers, R., Reuther, J., and Bange, F. C. (2009) The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of Mycobacterium tuberculosis. *Microbiology.* **155**, 1332–1339

175. Williams, K. J., Jenkins, V. A., Barton, G. R., Bryant, W. A., Krishnan, N., and Robertson, B. D. (2015) Deciphering the metabolic response of Mycobacterium tuberculosis to nitrogen stress. *Mol. Microbiol.* **97**, 1142–1157

176. Akhtar, S., Khan, A., Sohaskey, C. D., Jagannath, C., and Sarkar, D. (2013) Nitrite reductase NirBD is induced and plays an important role during in vitro dormancy of Mycobacterium tuberculosis. *J. Bacteriol.* **195**, 4592–4599

177. You, D., Xu, Y., Yin, B. C., and Ye, B. C. (2019) Nitrogen regulator GlnR controls redox sensing and lipids anabolism by directly activating the whiB3 in Mycobacterium smegmatis. *Front. Microbiol.* **10**, 1–10

178. Shukla, J., Gupta, R., Thakur, K. G., Gokhale, R., and Gopal, B. (2014) Structural basis for the redox sensitivity of the Mycobacterium tuberculosis SigK-RskA σ-anti-σ complex. *Acta Crystallogr. D. Biol. Crystallogr.* **70**, 1026–1036

179. Prusa, J., Zhu, D. X., and Stallings, C. L. (2018) The stringent response and Mycobacterium tuberculosis pathogenesis. *Pathog. Dis.* **76**, 1–13

180. Wu, X., Haakonsen, D. L., Sanderlin, A. G., Liu, Y. J., Shen, L., Zhuang, N., Laub, M. T., and Zhang, Y. (2018) Structural insights into the unique mechanism of transcription activation by Caulobacter crescentus GcrA. *Nucleic Acids Res.* **46**, 3245–3256