Summer 8-15-2016

Regulation of Transcription and Stress Response by CarD in Mycobacteria

Ashley Louise Garner
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Program in Molecular Microbiology and Microbial Pathogenesis

Dissertation Examination Committee:
Christina Stallings, Chair
Michael Caparon
Eric Galburt
Petra Levin
Jennifer Philips
Joseph Vogel

Regulation of Transcription and Stress Response by CarD in Mycobacteria

by
Ashley Louis Garner

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

August 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

The work in this thesis would not have been possible without the guidance, support and mentorship of many people. I am grateful to my thesis committee for their guidance, support, and advice. Their suggestions played a huge role in shaping this research. My thesis mentor Christina Stallings was indispensable for this project and her mentorship has played a crucial role in my maturation as a scientist. My time in the Stallings laboratory has been an extremely positive experience. It is a wonderfully supportive laboratory which has fostered amazing friendships. I would particularly like to thank Leslie Weiss, Kelly Flentie, Jacqueline Kimmey, and Katherine Mann who have been there throughout my thesis and have made the laboratory enjoyable in even the most trying times. Throughout my graduate studies I have appreciated the role the administrative and technical staff of the microbiology department for their role in managing such a well-run department which has facilitated many aspects of my studies. My friends, specifically Leah Imlay, Katrina Koc, and Mathew Kilgore have been a great support system and I wouldn't have made it through graduate school without time. Finally, I would like to thank my family: my parents Lee Bendig and Keith Garner, my siblings Graham and Nicole, and my sister-in-law Morgan. Your love and support have been so important to me.

Dedication

To my parents for their love and support

Abstract of the Dissertation

Regulation of Transcription and Stress Response by CarD in Mycobacteria

by

Ashley Louise Garner

Doctor of Philosophy in Biology and Biomedical Sciences

Program in Molecular Microbiology and Microbial Pathogenesis

Washington University in St. Louis, 2016

Professor Christina Stallings, Chair

Mycobacterium tuberculosis, the causative agent of Tuberculosis, infects over a third of the world's population. To control this epidemic, we must develop new chemotherapeutic treatments, which requires further insight into the physiology of this bacterium. Previous studies have identified CarD as a transcriptional regulator essential during both acute and persistent infection. Depletion of CarD sensitizes strains to a diverse panel of stresses and deregulates several hundred genes and ribosomal RNA (rRNA). Chromatin immunoprecipitation sequencing experiments showed that CarD was localized to promoters throughout the genome, suggesting that CarD regulates transcription initiation.

In collaboration with the Darst Lab, we published the first crystal structure of a CarD homolog. CarD's N-terminal domain was homologous to a known RNA polymerase (RNAP) interacting domain and C-terminal domain with a novel fold. Modeling CarD onto initiation structures of RNAP position CarD's C-terminal domain to interact with DNA. We identified three independent activities of CarD: binding RNAP, binding DNA, and the activity of a highly conserved tryptophan residue that we predict stabilizes the transcription bubble. Using a panel of

single mutations in *carD* that attenuate one of these three activities, I characterized the roles of each of CarD's activities *in vivo* and *in vitro*. All three of CarD's activities are necessary for optimal growth, antibiotic resistance, stabilizing RNAP-promoter complexes, and activating transcription from rRNA promoters. This work contributed to a model in which CarD slows the rate of transcription initiation DNA bubble collapse and accelerates DNA opening.

In further studies of CarD, I discovered a correlation between the cellular concentration of CarD and growth rate and showed that this growth rate dependence is not due to an effect on the rRNA content of the cell. This separated CarD's effect on growth rate from its effect on rRNA content for the first time, which indicates that this growth defect is a result of deregulation of non-rRNA promoters. Additionally, I elucidated a new mechanism of regulating CarD activity through turnover of free protein.

Most recently, I discovered the extent of CarD regulation in mycobacteria through RNA sequencing experiments. These studies revealed that more than 80% of the transcripts in the genome are significantly affected by alterations in CarD activity. Furthermore, there are transcript-dependent effects of CarD that imply that CarD is responsive qualities of the promoters but showed that the promoter sequence only partial explains this specificity.

My thesis work has dramatically advanced our understanding of the mechanism of transcriptional regulation by CarD. CarD's activity at transcription initiation complexes is an entirely novel mechanism of transcriptional regulation, creating a new paradigm of transcriptional regulation in prokaryotes. Furthermore, as CarD is conserved in many bacteria, its function has broad implications for bacterial transcription beyond mycobacteria. Finally, since CarD is essential in mycobacteria and absent from eukaryotes, my work will inform the development of new strategies to inhibit CarD activity as novel therapies to treat tuberculosis.

Chapter 1: Introduction

Ashley L. Garner, Kelly Flentie, and Christina L. Stallings

Portions of this document were originally published in (1).

Tuberculosis: epidemiology and disease progressions

Mycobacterium tuberculosis (*Mtb*), the causative agent of tuberculosis, is an obligate pathogen that evolved in Africa with early humans and has caused disease throughout history (2, 3). Today, about a third of the human population is infected *Mtb* and it causes more than a million deaths per year (4). The majority of infected individuals have latent tuberculosis, in which they are infected with live bacteria but have no symptoms and are not contagious. About 10% of immunocompetent patients with latent tuberculosis will progress to active disease over their lifetime but this rate increases to a 5-15% reactivation rate per year among HIV positive individuals (5, 6). Active tuberculosis is most commonly a pulmonary disease where it causes a serious cough that can produce blood in the sputum, chest pain, fever, weakness, loss of appetite, night sweats, and weight loss. The bacterium can also disseminate from the lungs to infect other body sites, including the brain, bones, or diffuse bacteremia. This disease presentation is most common in children. The Bacillus Calmette–Guérin (BCG) vaccine against *Mtb* was created in the 1920s and has been in common use since 1948. The vaccine effectively prevents disseminated disease, including meningitis, but does not protect against development or reactivation of pulmonary disease and therefore has been unable to eradicate the disease.

Treatment of active tuberculosis requires a two month intensive phase of treatment with ethambutol, isoniazid, pyrazinamide, and rifampicin, the four frontline antibiotics, followed by a four month continuous phase of treatment with isoniazid and rifampicin (7). The long treatment regimen, side effects of treatment, and limited access to healthcare in less developed countries all lead to poor patient compliance, which results in the development of antibiotic resistance. In 2014, 480,000 patients developed multidrug-resistant tuberculosis, which is resistant to at least two frontline antibiotics. Treatment of multidrug resistant tuberculosis requires treatment with at

least 4 antibiotics the bacterium is susceptible to for 6 months followed by an at least an 18 month long continuous treatment phase with multiple antibiotics (8). The limited antibiotics available to treat multidrug resistant tuberculosis infections and poor patient compliance results in a global cure rate of only 50%. Multidrug-resistant tuberculosis can evolve resistance to additional antibiotics resulting in about 10% of these infections being extensively drug-resistant and some strains gaining resistance to all currently antibiotics available to treat tuberculosis (4). Control of the global tuberculosis epidemic will require the development of novel antibiotics, development of which requires an enhanced understanding of mycobacterial physiology and virulence.

Infection with *Mtb* occurs when the bacterium is inhaled into the lungs where it is engulfed by a phagocytic cell of the innate immune system. This interaction initiates a pro-inflammatory response in which lymphocytes, monocytes, and fibroblasts are recruited to the site of infection and contain the infection by forming a granuloma (9). In most people, the immune system controls bacterial replication within the granuloma by exposing the bacterium to reactive oxygen and nitrogen species, attacking the mycobacterial cell surface with antimicrobial peptides, causing damage to the bacterium's DNA, and limiting the bacterium's access to oxygen, nutrients, and phosphate (10–21). While some patients who control infection likely eliminate the bacterium, others develop a latent infection in which bacterial loads are undetectable but remain present for the lifetime of the host (22). Other infected individuals, especially those who are immunocompromised, fail to control the infection and develop active disease. In these patients, the bacteria are not sealed within a granuloma and bacterial replication is not controlled. These individuals develop symptoms of infection, are contagious, and can die from the infection if it isn't successfully treated.

The most common animal model used to study tuberculosis infection is the mouse. In murine infection, the bacteria replicate for the first 2-3 weeks of infection after which bacterial burden is prevented from increasing by the onset of adaptive immunity. Like humans, the mouse's immune system imposes stresses on the bacteria and prevents replication, but differs from humans as mice do not form a true granuloma and the bacterial burden remains high for the lifetime of the mouse (23).

***Mtb*'s transcriptional response to stress**

During infection, the ability of the bacterium to survive the onslaught of stresses imposed by the host is essential for survival and therefore targeting cellular factors involved in stress response may be a promising avenue for development of novel therapeutics. In order to survive these stresses, *Mtb* executes a complex, interconnected web of stress responses that rely on changes in gene expression. The responses to different stresses are integrated and coordinated, often resulting in overlapping regulons and stress responders. Not only do these highly effective stress response strategies protect *Mtb* from host immunity, but the resulting changes in physiology also contribute to antibiotic tolerance, which precludes eradication of the infection (24–29). Coordinating stress responses, virulence, and antibiotic tolerance requires rigorous regulation of transcription. Transcription in all bacteria is achieved by a single core RNAP enzyme, consisting of the essential subunits β, β', and 2 α subunits along with the non-essential ω subunit (30, 31). To recognize and bind promoter sequences upstream from genes, the core RNAP associates with a σ subunit to form an RNAP holoenzyme. Most transcriptional regulation occurs at the level of initiation (32) and transcription factors can mediate this regulation by directly affecting the polymerase-promoter interaction, manipulating the equilibrium between

closed and open RNAP-promoter complexes (RP_c and RP_o respectively), or affecting rates of promoter escape (33, 34).

While many of the transcriptional response pathways are similar in *Mtb* and *Mycobacterium smegmatis*, a model organism for *Mtb*, there are differences between the two organisms in the number of a transcriptional regulators encoded in the genomes, the regulon of conserved regulators, and the essentiality of the regulators. *M. smegmatis* is a nonpathogenic, soil dwelling bacterium that replicates much faster than *Mtb* and must survive in a broader range of environments than the obligate pathogen *Mtb*. The more limited scope of environments *Mtb* is exposed to has allowed it to survive with a genome roughly two-thirds the size of *M. smegmatis*. The larger genome of *M. smegmatis* has allowed for an expansion is many transcription regulators as compared to *Mtb*. Despite these differences, many of the mechanisms by which the bacteria regulate transcription is similar in the two organism enabling many processes to be first studied in *M. smegmatis*. However, because there are differences between the bacteria, whenever possible researchers should verify that their findings are applicable to *Mtb* by directly testing hypotheses in *Mtb*.

Sigma factors globally control gene expression. The first determinant of gene expression in response to different conditions is the activity of the σ factor repertoire. Each σ factor binds a specific promoter sequence, thus conferring what promoters are targeted by the RNAP holoenzyme for transcription. Changes in σ factor activity in response to different stresses and conditions are able to shift a bacterium's expression profile. Compared to other obligate human pathogens, *Mtb* encodes the highest ratio of σ factors to genome size (35), which allow the bacterium to tailor its expression profile in response to a given environment. Even during exponential growth in culture, traditionally thought of as a relatively stress-free environment,

Mtb expresses its entire complement of σ factors (36–38), indicating that *Mtb* is poised to quickly respond to stress. *Mtb*'s σ factor network includes one essential housekeeping group 1 σ factor (σ^A), one stress-responsive group 2 σ factor (σ^B), and 11 group 3 and 4 alternative σ factors that also function as environmentally responsive regulators (σ^{C-M}) (35, 37, 39). This broad panel of σ factors allows *Mtb* to tune its transcriptional response for a large and diverse set of conditions. All of the σ factors in *Mtb* belong to the σ^{70} family, whose members in *E. coli* recognize two sequences in the promoter DNA, the -10 element (recognized by sigma region 2.4) and the -35 element (recognized by sigma region 4.2) (40). *Mtb* promoters contain a conserved -10 sequence that is essential and sometimes sufficient for transcription, while the -35 sequences are less conserved (40–42). The spacer region between -10 and -35 sequences in *Mtb* also varies dramatically compared to *E. coli* promoters (40, 43, 44). These differences in promoter elements may reflect the sigma diversity in *Mtb* (40, 44).

Evidence that alternative σ factors are important in *Mtb* during infection has come from cell culture and animal infection models. *sigE, sigF, sigG, sigH,* and *sigJ* are upregulated during infection of macrophages (45, 46) and both *sigE* and *sigG* are necessary for survival within macrophages (47–49). Deletion of *sigB*, *sigG, sigJ,* or *sigM* has no effect in animal models (37, 49–51). Deletion of *sigD, sigE, sigH,* or *sigL* results in delayed time to death without affecting bacterial burden (52–55), while deletion of *sigC* and *sigF* results in a delayed time to death and a decrease in bacterial burden during acute (*sigC*) or chronic (*sigF*) infection (56–58). The importance of individual σ factors during infection highlights both their central role in guiding *Mtb*'s stress response as well as the diverse adverse conditions encountered by *Mtb* during infection.

The stringent response controls energy expenditure during stress. In addition to sigma factors, the stringent response is another conserved global stress response in bacteria that regulates growth and metabolism based on cellular energy levels in harsh environments. In mycobacteria, the stringent response is best characterized during amino acid starvation when the Rel_{Mtb} enzyme senses uncharged tRNAs in ribosomes and responds by transferring the pyrophosphate (PPi) group from ATP to GDP and GTP to synthesize hyperphosphorylated guanine nucleotides ppGpp and pppGpp (collectively called (p)ppGpp) (59). (p)ppGpp then coordinates downstream regulation of bacterial physiology and mediates changes in the transcriptional profile to support survival during stress. Deletion of rel_{Mtb} led to 159 genes differentially expressed during starvation, including genes involved in coordinating metabolic rate reduction, production of mycobacterial cell wall and lipids, secreted proteins, and cell division machinery (60). Rel_{Mtb} is required for survival in low nutrient conditions, long term culture, and during infection in animal models, all indicative of a strict requirement for Rel_{Mtb} during exposure to stress (60–64). In *E. coli*, (p)ppGpp directly affects transcription initiation by binding the RNAP (65, 66). In contrast, in a number of Gram-positive bacteria, (p)ppGpp inhibits GTP biosynthesis by directly interacting with GTP synthesis enzymes, which impacts gene expression by altering initiating nucleotide levels (66–69). Although (p)ppGpp has not been demonstrated to directly bind *Mtb* RNAP or GTP synthesis enzymes, (p)ppGpp has been reported to influence mycobacterial RNAP activity *in vitro*, suggesting that the mechanism of (p)ppGpp action in *Mtb* transcriptional modulation requires further investigation (65, 67, 70).

Two component systems control specific responses to stress. The transcriptional response to stress is further modified through regulation by two component systems and transcription factors that control expression of a limited set of genes in response to a specific

stress. By activating factors that regulate a smaller number of genes *Mtb* can tailor its transcriptional response to the specific environment it is encountering, which is particularly important for survival within the host. *Mtb* encodes 2 essential and 10 nonessential two-component systems, which are classically recognized as bacterial systems to sense and respond to stress and changes in the environment (71). Each two component system consists of at least one sensor histidine kinase that responds to specific environmental conditions by autophosphorylation and phosphotransfer to its cognate response regulator, which then binds DNA and activates transcription of a specific regulon (71).

In addition to its two component systems, transcription factors allow *Mtb* to respond to a stimulus through transcriptional expression or repression. Recently, researchers overexpressed 200 predicted TFs in *Mtb* and performed chromatin immunoprecipitation sequencing experiments and microarray analyses to catalogue a genome-wide characterization of TF binding events and target gene expression (72–74). These reports describe 16,000 binding sites for 154 TFs and identify regulatory routes for ~70% of the genome. The complex regulatory circuits that were uncovered highlight how much is still left to be investigated regarding how *Mtb* regulates transcription to integrate precise stress responses. While most transcription factors in *Mtb* are not essential, some iron binding transcription factors involved in regulating the cell's redox state or iron metabolism are essential. The essentiality of these factors, including *whiB1*, *whiB2*, and *ideR*, indicates a particular need for *Mtb* to couple redox sensing and iron availability with basic cellular processes to maintain homeostasis.

Additional determinants of stress response. The transcriptional regulators described above are in no way exhaustive in terms of all of the mechanisms of transcriptional regulation that *Mtb* employs to respond to stress. In particular, there is a growing area of research into the

roles of nucleoid associated proteins and small RNAs (75–78). *Mtb* also contains 11 serine/threonine protein kinases that, like two component systems, are involved in signal transduction pathways that aid *Mtb* in adaptation to its environment (79). However, unlike two component systems that consist of sensor kinases that activate response regulators to directly modulate *Mtb* transcription, serine/threonine protein kinases are single proteins that phosphorylate numerous downstream targets (79). Although serine/threonine protein kinases do not directly affect *Mtb* transcription, they do influence gene expression by modifying the activity of other *Mtb* proteins with more direct roles in transcription, such as sigma factors, nucleoid-associated proteins, anti-anti-sigma factors, and two component systems (80–84).These and other aspects of gene regulation further add to the complexity of stress responses in *Mtb*.

RbpA and CarD: stress responsive transcriptional regulators in actinobacteria.

Two important stress response regulators in mycobacteria that are not conserved across bacterial species are RbpA and CarD. Both of these proteins are RNAP-binding proteins that further modify gene expression from a given holoenzyme. While RbpA is limited to actinobacteria, CarD is present in numerous other bacterial phyla (85–87), including *Bacillus* and *Thermus*, but not in *E. coli. rbpA* is upregulated during oxidative stress, stationary phase, starvation, hypoxia, high temperatures, treatment with antibiotics, and during infection in macrophages (88–92). RbpA consists of a central RbpA core domain (RCD) flanked by an unstructured 26-aa N-terminal tail and a C-terminal σ interaction domain (SID) linked to the RCD by a 15-aa basic linker (BL) (93–95). RbpA forms a stable binary complex with the σ2-domain of group 1 (σ^A in *Mtb*) and certain group 2 σ factors (σ^B in *Mtb*) through its SID (93–95), with additional contacts made between the N-terminus and the σ-factor (93). Based on structural modeling, the RbpA BL domain and adjacent residues interact with the DNA phosphate

backbone of the nontemplate strand upstream of the -10 promoter element in the RP_o conformation (93). Additional contacts between RbpA and RNAP σ have been proposed based on crosslinking experiments (96–98), but the recent structural modeling of RbpA onto an RNAP-promoter open complex would be incompatible with these interactions (94), suggesting that further analysis will be needed to resolve these inconsistencies. RbpA has been shown to increase the affinity of the σ -factor to the core RNAP, increase the affinity of RNAP holoenzyme to promoter DNA, and facilitate the formation of RP_o (94, 99, 100), all of which could contribute to the ability of RbpA to promote RNAP-promoter complex formation and stability. The housekeeping σ factor, σ^A, has been reported to have a similar affinity for *Mtb* RNAP core enzyme as the alternative σ factor σ^F (98), in which case RbpA may be necessary to improve σ^A affinity and competitiveness for RNAP in conditions that require the activity of σ^A. In contrast, in *E. coli* σ^{70} has a very high affinity to the RNAP core enzyme and thus can out-compete other σ factors in conditions where it is required without accessary factors like RbpA.

CarD was initially identified as being upregulated in response to various stresses, including oxidative stress, starvation, and a broad panel of antibiotics, but is essential in *Mtb* even during growth in nutrient rich cultures (85). *In silico* analysis suggested that CarD's N-terminus is homologous to the RNAP Interaction Domain (RID) of the Transcription Repair Coupling Factor (TRCF) that is known to interact with the RNAP β-subunit β1-lobe. To test if CarD also interacts with RNAP, immunoprecipitation experiments in *M, smegmatis* were designed (85). Using the known interaction of TRCF with RNAP, a model of CarD's RID domain interacting with RNAP was constructed. This model predicted that an interaction between glutamate 138 (E138) of RNAP-β with arginines 25 and 47 (R25 and R47) of CarD was important for the strength of this complex (Fig 1A, (101)). To test the accuracy of this modeled

interaction, both arginines in CarD were separately mutated to glutamates (R25E and R47E) in *M. smegmatis* and the effect on the strength of the interaction was tested by an immunoprecipitation. Both mutations were found to decreased the affinity of the interaction, although CarD^{R25E} more dramatically attenuated the interaction than CarD^{R47E} (Fig 1B). Interestingly, while strains singly expressing either mutant were viable in *M. smegmatis*, only the CarD^{R47E} strain was viable in *Mtb*, suggesting that *Mtb* may have more stringent requirements for CarD than *M. smegmatis.* The less stringent requirements for CarD in *M. smegmatis* allow for a greater variety of genetic tools to be used in this organism than are viable in Mtb, thus representing a powerful model to study CarD's function. Mutations that decreased CarD's affinity for RNAP caused slower growth (Fig. 1C) and increased sensitivity to oxidative (Fig. 1D) stress as well as antibiotics (Fig. 1E). In *Mtb*, strains expressing CarD^{R47E} had lower bacterial burden in mice during persistent infections (Fig. 1FG). Together these data indicated that CarD's interaction with RNAP was crucial for CarD to support growth and stress resistance in mycobacteria (101).

Since CarD was shown to interact with RNAP, it was hypothesized that CarD may regulate transcription. To determine if CarD affects transcription, a microarray was performed comparing CarD-depleted cells to cells expressing CarD. Depletion of CarD resulted in 193 transcripts being deregulated greater than 2-fold suggesting that CarD was a transcriptional regulator. Genes involved in the translational machinery were over-represented among the deregulated transcripts and further analysis by quantitative reverse transcription PCR (qRT-PCR) found that depletion of CarD also significantly deregulated 16s ribosomal RNA (rRNA, (85)). These findings supported a role for CarD as a transcriptional regulator and specifically identified rRNA as a transcript dependent on CarD mediated regulation.

To determine which stage of transcription (initiation, elongation, termination) is regulated by CarD, chromatin immunoprecipitation sequencing (ChIP-seq) reactions were performed to determine the chromosomal localizations of a hemagglutinin (HA) tagged CarD, RNAP-β, and RNAP-σ^A (the housekeeping sigma factor) in *M. smegmatis*. RNAP-β was localized to DNA throughout the length of transcribed genes while RNAP-σ^A was specifically localized to promoter regions where transcripts were initiated. CarD was never localized to DNA in the absence of RNAP and its genomic localization was highly correlated to that of RNAP-σ^A (Fig 2). CarD was localized to promoter regions throughout the chromosome, including numerous promoter regions that were not deregulated by depletion of CarD (85, 87). The co-localization of CarD and RNAP-σ^A suggested that CarD was likely regulating transcription initiation and that it may be a more universal transcriptional regulator in mycobacteria than was suggested by the microarray of CarD depleted cells.

A major advance in CarD research occurred in 2013 when the crystal structure of the *Thermus thermophilus* CarD was published (87). Since then numerous other crystal and NMR structures of CarD homologs have been solved (86, 102–104). These structures confirmed the expected structure of the N-terminal RID domain and revealed a largely α-helical C-terminus that was not homologous to any known protein structure (Fig. 3A). Based on the known interaction between TRCF and RNAP-β, CarD was modeled onto RNAP structures formed during transcription initiation of the *Thermus aquaticus* RNAP (Fig. 3B). CarD was specifically modeled onto RNAP initiation complexes RP_c and RP_o as the ChIP-seq experiments had shown that CarD was localized to promoter regions. These models positioned CarD's C-terminal domain to interact with DNA just upstream of the promoter -10 motif via a conserved basic patch in CarD (Fig. 3C). Within this basic patch, there is a nearly universally conserved tryptophan

residue that sterically clashes with the DNA backbone of RP_c but not RP_o RNAP structures. CarD's ability to bind DNA was tested via an electrophoretic mobility shift assay (EMSA), which demonstrated that addition of CarD retarded the migration of a DNA fragment encoding the promoter and leader sequences of a ribosomal RNA operon from *M. smegmatis* (Fig. 3D, (87)). As a result of these experiments, it was found that CarD binds DNA via its C-terminal domain in a sequence independent manner. CarD's DNA binding ability and the structural studies suggested that CarD may regulate transcription by affecting RP_o formation or stability by a mechanism that requires contact with both RNAP and promoter DNA and paved the way for mechanistic studies to determine the effect of CarD on transcription.

Figure Legends

Figure 1: CarD's N-terminal interaction with RNAP is important for growth, stress resistance and virulence in *Mtb*

A. Homology model (105) of *Mtb* CarD-RID with *Mtb* RNAP-β1 based on the crystal structure of the *T. thermophilus* TRCF-RID with *T. aquaticus* RNAP-β1 (106). Proteins are shown as backbone ribbons, with the CarD-RID in light pink and RNAP-β1 in light cyan. Side chains of residues at the interaction interface discussed in the text are shown (CarD-RID, magenta; RNAP-β1, cyan).

B. Immunoprecipitation experiments with a monoclonal antibody specific for HA in *M. smegmatis ΔcarD attb*::tet*carD* expressing untagged $CarD^{WT}$ (lane 1, also expresses HA peptide), $CarD^{WT}$-HA (lane 2), $CarD^{R25E}$-HA (lane 3) or $CarD^{R47E}$-HA (lane 4). Inputs (before immunoprecipitation) and eluates were analyzed by western blotting with antibodies specific for either the RNAP β subunit or HA.

C. Representative growth curve of *Mtb ΔcarD attb*::tet*carD* expressing $CarD^{WT}$ or $CarD^{R47E}$.

D-E. Log phase *Mtb ΔcarD attb*::tet*carD* expressing $CarD^{WT}$ or $CarD^{R47E}$ growing in 7H9 broth were treated for 75 hours with (D) 25mM H_2O_2 or (E) 200 μg/ml rifampicin. After treatment, dilutions were plated on 7H10 and survival is graphically represented as the ratio of CFUs in treated as compared to untreated cultures.

F-G. Bacterial titers in the lungs (F) and spleens (G) of C57Bl/6 infected with *Mtb ΔcarD attb*::tet*carD* expressing $CarD^{WT}$ (open circles) or $CarD^{R47E}$ (open triangles).

Graphs in C-G show the mean ± SEM and each sample. Significance was determined by calculating the p value 724 from a student's t test where * is significant with a p value < 0.05, ** is significant 725 with a p value < 0.01, and *** is significant with a p value <0.005.

All panels in this figure were first published in (101).

Figure 2: Normalized log2 of ChIP-seq reads from *M. smegmatis* DNA coimmunoprecipitated with RNAP β, σ^A, or CarD-HA. Aggregate profiles averaged over 62 highly active transcription units. Protein–DNA complexes containing CarD-HA, RNAP β, and RNAP σ were immunoprecipitated from *M. smegmatis* lysates. The coprecipitated DNA was sequenced, and the number of sequence reads per base pair was normalized to total reads per sample and expressed as a $\log_2$ value. Normalized reads per base pair from DNA precipitated from cells expressing only the HA epitope were used as background and subtracted from the other samples. The 62 transcription units were selected on the basis of high signal and isolation from sur- rounding transcription units. Data presented in this figure was published in (87).

Figure 3: Primary and 3D structural features of *Mtb* CarD.

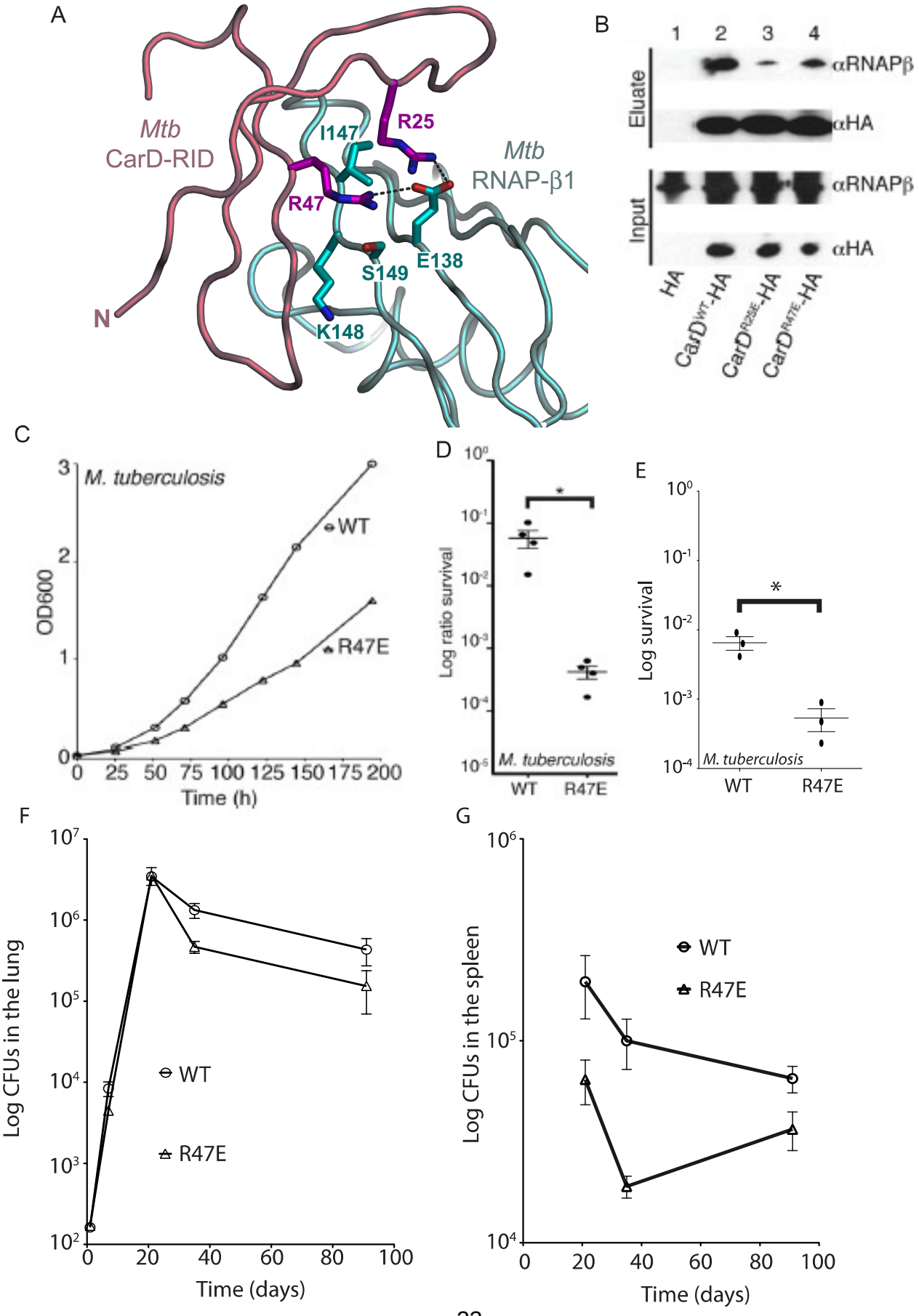
A. Crystal structure of *T. thermophilus* CarD (87), shown in ribbon format. The N-terminal CarD-RID domain is shown in magenta and the C-terminal domain is in green.

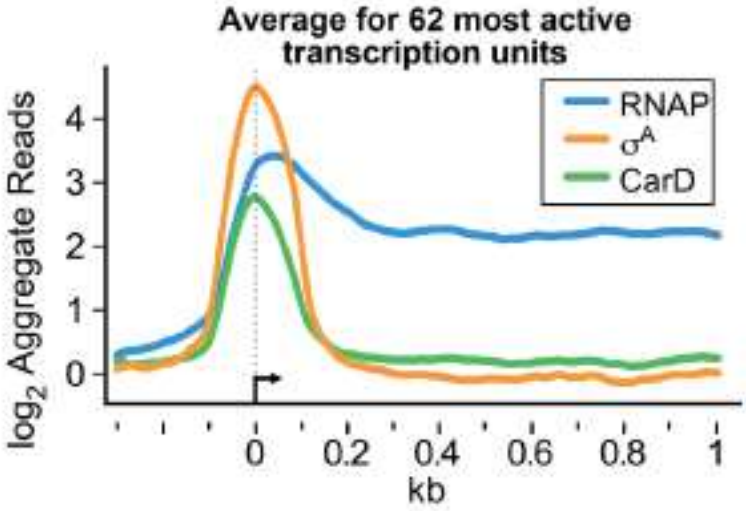
B. Structural model of the *Thermus* CarD/RP_o complex. The model was generated as described in (87). The colors correspond to: CarD RID is in magenta, CarD C-term is in green, RNAP-β is in blue, RNAP-β' is in pink, RNAP-σ is in orange, DNA is in green with the promoter element in yellow

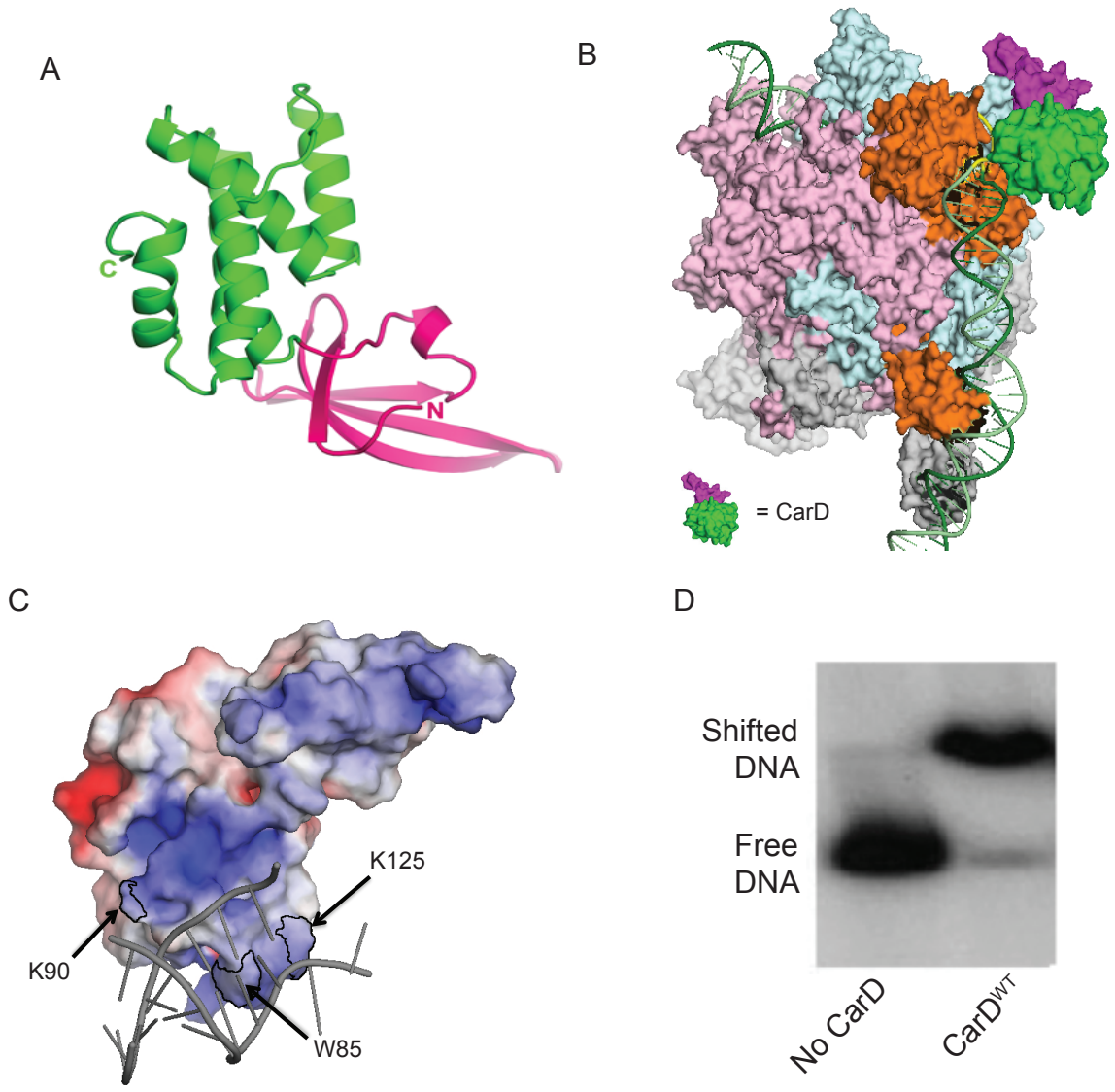
C. Crystal structure of *T. thermophilus* CarD, shown as a molecular surface and colored according to the electrostatic surface potential [red, −5 kT; white, neutral; blue, +5 kT; where k is the Boltzmann constant, T is temperature (107). Three conserved residues in *Mtb* and their locations are indicated on the structure.

D. Autoradiographs of EMSAs with 200 pmol CarD or no protein incubated with 20,000 cpm (~0.6 ng) of [γ32P]ATP-radiolabeled *rrnA*PL DNA. The reactions were separated on a nondenaturing polyacrylamide gel, which was then dried and exposed to film.

Data presented in this figure was published in (87).


A
Mtb
CarD-RID
Mtb
RNAP-β1
I147
R25
R47
S149
E138
K148
N
B
1
2
3
4
Eluate
αRNAPβ
αHA
Input
αRNAPβ
αHA
HA
CarD^WT-HA
CarD^R25E-HA
CarD^R47E-HA
C
M. tuberculosis
WT
R47E
OD600
Time (h)
D
Log ratio survival
M. tuberculosis
WT
R47E
E
Log survival
M. tuberculosis
WT
R47E
F
Log CFUs in the lung
WT
R47E
Time (days)
G
Log CFUs in the spleen
WT
R47E
Time (days)


Average for 62 most active transcription units
RNAP
σA
CarD
log2 Aggregate Reads
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1
kb


A
C
N
B
= CarD
C
K125
K90
W85
D
Shifted
DNA
Free
DNA
No CarD
CarD^WT

Chapter 2: CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria

Ashley L. Garner, Leslie A. Weiss, Ana Ruiz Manzano, Eric A. Galburt, and Christina L. Stallings

A version of this chapter was originally published as (108). This chapter also details findings originally published in (87, 109)

Abstract

Although the basic mechanisms of prokaryotic transcription are conserved, it has become evident that some bacteria require additional factors to allow for efficient gene transcription. CarD is an RNA polymerase (RNAP) binding protein conserved in numerous bacterial species and essential in mycobacteria. Despite the importance of CarD, its function at transcription complexes remains unclear. We have generated a panel of mutations that individually target three independent functional modules of CarD: the RNAP interaction domain, the DNA binding domain, and a conserved tryptophan residue. Using these mutants and a combination of *in vivo* and *in vitro* approaches, we have discovered that CarD activates transcription by stabilizing an RNAP-promoter complex at ribosomal RNA promoters. We have dissected the roles of each functional module in CarD activity and built a model where each module contributes to stabilizing RNAP-promoter complexes. Our work highlights the requirement of all three modules of CarD in the obligate pathogen *Mycobacterium tuberculosis* (*Mtb*), but not in *Mycobacterium smegmatis*. We also report divergent use of the CarD functional modules in resisting oxidative stress and pigmentation. These studies provide new information regarding the functional domains involved in transcriptional regulation by CarD while also improving understanding of the physiology of *Mtb*.

Introduction

CarD is an essential transcriptional regulator in mycobacteria. Its N-terminal domain is homologous to the RNAP interacting domain (RID) of the transcription repair coupling factor (TRCF, encoded by *mfd*). *In vivo* studies have characterized CarD mutants with decreased affinity for RNAP and shown that weakening this interaction decreases growth rate, increases stress sensitivity, and decreases virulence of *Mtb*, demonstrating this importance of this

interaction for CarD's function (Chapter 1 Fig. 1, (101)). Evidence that CarD regulates transcription comes from a microarray study performed in Mycobacterium smegmatis that found deregulation of several hundred genes, including many genes associated with the translational machinery, upon depletion of the CarD protein (85).

To further our understanding of CarD, the crystal structure of the *Thermus thermophilus* CarD was recently solved. This structure confirmed the predicted structure of the N-terminal domain and revealed the C-terminal domain to be largely alpha-helical and not homologous to any previously solved protein domain (Chapter 1 Fig. 3, (87)). In an effort to determine what stage of transcription may be regulated by CarD, chromatin immunoprecipitation sequencing experiments were performed. These studies demonstrated that CarD is localized with RNA polymerase (RNAP) holoenzyme to DNA at promoter regions throughout the genome and has the same genomic distribution as the house keeping sigma factor, which supported a role for CarD in regulation of transcription initiation (Chapter 1 Fig. 2, (87)). Because CarD was proposed to regulate transcription initiation, CarD was modeled onto initiation complexes of the *Thermus aquaticus* RNAP. This model positions the C-terminal domain to interact with DNA just upstream of the -10 element of the promoter via a conserved basic patch within which is a highly conserved tryptophan residue. CarD's ability to bind DNA was tested by an electrophoretic mobility shift assay that demonstrated that CarD binds DNA without sequence specificity (Chapter 1 Fig. 3, (87)).

Numerous questions remain as to the mechanism by which CarD regulates transcription and the importance of CarD's interaction with DNA. In this chapter we demonstrate for the first time both that CarD regulates transcription from promoters *in vivo* and that it stabilizes a RNAP-promoter complex during transcription initiation. Furthermore, we investigate many open

questions including what residues within CarD are necessary for its interaction with DNA, what the roles of this interaction are in CarD-mediated transcriptional regulation and pathogenesis, and how do the CarD/RNAP and CarD/DNA interactions affect one another. To address these questions, we isolated single point mutants in the CarD C-terminus and use these mutants to parse out the role of the CarD C-terminus in DNA binding, bacterial stress responses, pathogenesis, and regulation of transcription. In addition, we have begun to clarify the relationship between the RNAP binding and DNA binding activities of CarD. Our studies demonstrate that a conserved tryptophan residue in CarD that is physically located within the DNA binding domain does not phenotypically group with the mutants that abolish CarD's interaction with DNA and performs a function that is important for CarD activity but is distinct from the DNA and RNAP binding activities. This is the first report to mechanistically investigate transcriptional regulation by CarD and to analyze the CarD C-terminal mutants *in vivo*. These studies reveal a functional distinction between the conserved tryptophan and the surrounding basic patch in the C-terminus of CarD. Together our studies dissect three functional regions of CarD and elucidate their interplay as it relates to CarD activity. These studies represent a step in understanding the mechanism of transcriptional regulation by CarD.

Experimental Procedures

Media and bacterial strains. (i) *Mtb.* All *Mtb* strains were derived from the Erdman strain and were grown at 37°C in 7H9 (broth) or 7H10 (agar) (Difco) media supplemented with 60 µl L^{-1} oleic acid, 5 g L^{-1} bovine serum albumin (BSA), 2 g L^{-1} dextrose, 0.003 g L^{-1} catalase (OADC), 0.5% glycerol, and 0.05% Tween 80 (broth). Gene switching was used to construct strains of mycobacteria expressing different *carD* alleles and to test for their viability (110). Specifically, the *Mtb* Δ*carD attB*::*tet-carD* strain (described previously in (85)) was transformed

with pMSG430*smcarD*, pMSG430*smcarD*^{W85A}, pMSG430*smcarD*^{K90A}, pMSG430*smcarD*^{K90E}, pMSG430*smcarD*^{K125A}, or pMSG430*smcarD*^{K125E} (expresses *M. smegmatis* CarD^{WT}, CarD^{W85A}, CarD^{K90A}, CarD^{K90E}, CarD^{K125A}, or CarD^{K125E}, respectively, from a constitutive P*myc1-tetO* promoter, kanamycin resistant) to replace the pDB19-*Rv3583c* construct (expresses *Mtb* CarD from a constitutive P*myc1-tetO* promoter, zeocin resistant) at the *attB* site of *Mtb* Δ*carD attB*::*tet-carD*. The transformants were selected on kanamycin. The *carD* gene from each transformant was sequenced to confirm the presence of the correct sequence. The *Mtb* Δ*carD attB*::*tet-carD* strains transformed with pMSG430*smcarD* and pMSG430*smcarD*^{K125A} were named csm41 and csm45, respectively. Previously described *Mtb* Δ*carD attB*::*tet-carD* strains transformed with pDB19-*Rv3583c*^{WT} and pDB19-*Rv3583c*^{R47E}, named mgm3080 and mgm3081 respectively, were also used in this paper (101). The *Mtb* and *M. smegmatis* CarDs have only 3 conservative amino acid differences and complement each other both *in vivo* and *in vitro*.

(ii) *M. smegmatis*. All *M. smegmatis* strains were derived from mc^{2}155 and were grown at 37°C in LB supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% Tween 80 (broth). The *M. smegmatis* strains expressing either hemagglutinin (HA)-tagged or untagged CarD^{WT}, CarD^{R25E}, CarD^{R47E}, CarD^{W85A}, CarD^{K90A}, CarD^{K90E}, CarD^{K125A}, or CarD^{K125E} were engineered as described for the analogous *Mtb* strains using pMSG430 expression plasmids and the *M. smegmatis* Δ*carD attB::tet-carD* strain (described previously in (85)). The *M. smegmatis* Δ*carD attB*::*tet-carD* strains expressing *Mtb* CarD^{WT}, *Mtb* CarD^{R25E}, *Mtb* CarD^{R47E}, *M. smegmatis* CarD^{W85A}, *M. smegmatis* CarD^{K90A}, *M. smegmatis* CarD^{K90E}, *M. smegmatis* CarD^{K125A}, or *M. smegmatis* CarD^{K125E} from a constitutive P*myc1-tetO* promoter at the *attB* site of *M. smegmatis* Δ*carD attB*::*tet-carD* were named mgm3043, mgm3044, mgm3045, csm34, csm32, csm33, csm35, and csm36, respectively. The strains expressing these same alleles but as C-terminal HA-

tagged versions of CarD were named mgm3090, mgm3091, mgm3092, csm50, csm48, csm49, csm51, and csm52, respectively.

All *Mtb* and *M. smegmatis* strains used in this chapter contain only one *carD* allele.

Antibiotics and chemicals. In mycobacterial cultures, 20 μg ml^{-1} kanamycin, 12.5 μg ml^{-1} zeocin, 20 μg ml^{-1} streptomycin, 50 μg ml^{-1} hygromycin, and 50 ng ml^{-1} of anhydrotetracycline (ATc) were used. H_2O_2 (Fisher Scientific) was used at 25mM.

Native Gel Electrophoresis Mobility Shift Assays (EMSAs). A DNA fragment containing the *M. smegmatis rrnA* promoter and leader sequences, called *rrnA*PL and corresponding to *M. smegmatis* mc^2155 nucleotides 5,029,577–5,029,909, was used for EMSAs as previously described (Srivastava et al., 2013). The DNA (250 ng) was amplified with IRDye labelled primers. 1 pmol of labelled DNA, 1 μg of LightShift Poly(di/dc) competitor DNA (Thermo Scientific), 10 μg BSA, and 200 pmol of CarD proteins were mixed with binding buffer (20 mM Tris, pH 8, 150 mM NaCl) in a total volume of 20 μl and incubated for 20 min at room temperature. Samples were then electrophoresed on 4–20% non-denaturing TBE polyacrylamide gels (Invitrogen) and imaged using an Odyssey CLX imaging system (LI-COR).

Western blotting and immunoprecipitation. For immunoprecipitation, 50 ml cultures were washed and lysed in 500 μl of NP-40 buffer (10 mM sodium phosphate, pH 8.0, 150 mM NaCl, 1% Nonidet® P-40, and Roche Complete protease inhibitor cocktail) by bead beating (FastPrep; MP Bio). Twenty-five microliters of lysate was used for the input sample and the rest was treated with DNase I (New England BioLabs), added to monoclonal anti-HA agarose (Sigma), and rotated overnight at 4°C. The matrix was washed 3 times with NP-40 buffer, and immunoprecipitated protein complexes were eluted with 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 500 μg ml^{-1} HA peptide (Roche), and protease inhibitors. For the western blot analyses, CarD-

HA and RNAP β were detected using mouse monoclonal antibodies specific for CarD (clone 10F05; Memorial Sloan-Kettering Cancer Center Monoclonal Antibody Core Facility) and RNAP β (clone 8RB13; Neoclone, Madison, WI), respectively.

Animal infections. Before infection, exponentially replicating *Mtb* csm41 ($CarD^{WT}$) and csm45 ($CarD^{K125A}$) strains were washed in PBS plus 0.05% Tween-80 and sonicated to disperse clumps. Eight- to nine-week-old female C57BL/6 mice (Jackson Laboratory) were exposed to 8×10^7 CFU of the appropriate strain in an inhalation exposure system (Glas-Col) that delivers ~100 bacteria per animal. The bacterial burden was determined by plating serial dilutions of lung and spleen homogenates onto 7H10 agar plates. The plates were incubated at 37°C in 5% CO_2 for 3 weeks prior to counting colonies. All procedures involving animals were conducted according to the National Institutes of Health (NIH) guidelines for the housing and care of laboratory animals, and they were performed in accordance with institutional regulations after protocol review and approval by the Institutional Animal Care and Use Committee of the Washington University in St. Louis School of Medicine (protocol 20130156, Analysis of Mycobacterial Pathogenesis). Washington University is registered as a research facility with the United States Department of Agriculture and is fully accredited by the American Association of Accreditation of Laboratory Animal Care. The Animal Welfare Assurance documentation is on file with the Office for Protection from Research Risks of the NIH. All animals used in these experiments were subjected to no or minimal discomfort. All mice were euthanized by CO_2 asphyxiation, which is approved by the American Veterinary Medical Association Panel on Euthanasia.

Survival assays. Zone of inhibition assays: 500 µl of a log-phase *M. smegmatis* culture was plated on LB agar supplemented with 0.5% dextrose and 0.5% glycerol. A single 6-mm disk

(Sigma) was placed in the middle of the freshly plated bacterial lawn, and 5 µl of 1 mg ml^{-1} ciprofloxacin, 100 mg ml^{-1} rifampicin, or 200 mg ml^{-1} streptomycin was spotted on the disk. The plates were then incubated for 2 days at 37 ºC before the radius of the zone of inhibition was measured. For transient treatment in liquid culture assays, log-phase *M. smegmatis* cultures in LB broth supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% Tween 80 were treated with 10 µg ml^{-1} ciprofloxacin for 2 hrs or 25 mM H_2O_2 for 1 hr before plating dilutions made directly from the treated cultures.

qRT-PCR. RNA was prepared from 5–10 mL of log-phase *M. smegmatis* mgm3043 ($CarD^{WT}$), mgm3044 ($CarD^{R25E}$), mgm3045 ($CarD^{R47E}$), csm32 ($carD^{K90A}$), csm34 ($CarD^{W85A}$), csm35 ($CarD^{K125A}$), and csm36 ($CarD^{K125E}$) or *Mtb* csm41 ($CarD^{WT}$) and csm45 ($CarD^{K125A}$) and 16S rRNA levels were measured and normalized to *sigA* transcript levels as previously described (85).

β-galactosidase Assays. *M. smegmatis* mgm3043, mgm3044, mgm3045, csm32, csm34, csm35, and csm36 were transformed with pHMG147-AP1-*lacZ*, pHMG147-AP2-*lacZ*, or pHMG147-AP3-*lacZ* (HygR episomal plasmid that expresses *lacZ* from the indicated promoters) to perform β-galactosidase assays in strains expressing CarD mutants. β-galactosidase activity was measured from log-phase cultures by pelleting the cells and washing them in Z buffer (60 mM Na_2HPO_4, 60 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7) before resuspending them in 200 µL of Z buffer and bead beating four times (FastPrep; MP Bio) to lyse the cells. Total protein content in each lysate was measured by BCA assay (Pearce). A 50 mM final concentration of BME was added to the remaining lysate preparation and β-galactosidase activity was measured by adding a final concentration of 2 mg mL^{-1} ortho-nitrophenyl-β-galactoside (ONPG) solution in Z buffer. Change in $OD\lambda_{420}$ absorbance was measured over time using the

Epoch Microplate Spectrophotometer (BioTek). β-galactosidase activity was calculated using the equation β-galactosidase activity = V max /(total protein concentration × volume of lysate) with V max being the rate of change in ODλ_{420} absorbance (measured as the slope).

***In Vitro* Transcription Assays.** The *rrnA*P123 (nucleotides 5,029,667-5,029,905 of the *M. smegmatis* mc2155 genome), *rrnA*P13 (nucleotides 1,469,982-1,470,234 of the *Mtb* Erdman strain), and *rrnA*P3 (nucleotides 1,470,113-1,470,157 of the *Mtb* Erdman strain) promoter fragments used in the *in vitro* transcription assays were cloned into the EcoRI and HindIII restriction enzyme recognition sites in the pRLG770 plasmid, which has been used previously for similar assays (111). The DNA segment from pRLG770 (112) containing the promoter, test transcript template, and terminator was then cloned into the NcoI and PstI sites in the pGem-T (Promega) plasmid. The plasmids were prepared from *E. coli* by midiprep (Qiagen) and phenol/chloroform extraction. CarD proteins used in *in vitro* transcription assays were diluted into 1 × dialysis buffer (20 mM Tris pH8.0, 150 mM NaCl, and 1 mM BME). Recombinant *M. bovis* core RNAP was purified from *E. coli* using a system kindly supplied by Dr. Robert Landick (113, 114). Recombinant *M. bovis* σ^A was also purified from *E. coli* and added to the core RNAP to reconstitute the RNAP holoenzyme. For complex stability assays (87), 200 nM (for *rrnA*P3 or *rrnA*P13 assays) or 400 nM (*rrnA*P123 assays) *M. smegmatis* CarD^{WT}, *Mtb* CarD^{R25E}, *M. smegmatis* CarD^{W85A}, *M. smegmatis* CarD^{K90A}, or *M. smegmatis* CarD^{K125E} was preincubated with 20 nM (for *rrnA*P3 or *rrnA*P13 assays) or 40 nM (*rrnA*P123 assays) *Mycobacterium bovis* core RNAP in 1x storage buffer (10 mM Tris·HCl, pH 8.0, 3 mM $MgCl_2$, 50 mM NaCl, 1 mM BME, and 50% (vol/vol) glycerol) for 10 min on ice, followed by the addition of 40 nM (for *rrnA*P3 or *rrnA*P13 assays) or 80 nM (*rrnA*P123 assays) *M. bovis* σ^A (in 1x storage buffer). After 10 min, 50 ng of a supercoiled plasmid DNA containing a test promoter

was added and the reaction was brought to 12.5μL by dilution such that the final solutions contained 1 × transcription buffer (10 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 50 μg mL^{-1} BSA, 40 mM NaCl, and 1 mM DTT). 400 nM competitor DNA (double-stranded FullCon promoter DNA fragment (111)) was added and at the designated times after the addition of competitor, transcription was initiated by the addition of NTPs (100 μM GTP, CTP, and ATP; 10 μM UTP; and 0.1 μL [α -^{32}P]UTP for *rrnA*P3 and *rrnA*P13 or 100 μM CTP and ATP, 500 μM GTP, 10 μM UTP, and 0.1 μL [α -^{32}P]UTP for *rrnA*P123). A higher concentration of the initiating nucleotide GTP is required to achieve enough transcript from the *M. smegmatis rrnA*P3 promoter to be reliably detectable in these assays. After 15 min, the reactions were stopped with 2 × formamide buffer (98% (vol/vol) formamide, 5 mM EDTA) and run on a 6% urea PAGE gel.

Results

Single point mutations in the CarD C-terminal basic patch abolish the interaction between CarD and DNA *in vitro*. To determine the residues in the CarD C-terminus that were responsible for the interaction with DNA, we purified CarD proteins harboring a single mutation in the tryptophan or lysine residues (W85, K90, and K125 (Fig. 1A)) that are highly conserved among CarD homologs (87). EMSAs were performed to assay the effect of mutating W85, K90, and K125 on the ability of CarD to bind the DNA fragment *rrnA*PL, which contains the promoters and leader sequences of the *M. smegmatis* ribosomal RNA (rRNA) *rrnA* operon. This is the same DNA fragment that was used in the experiments that identified the C-terminus of CarD as a DNA binding domain (87) and was originally chosen due to CarD's described role in regulating rRNA transcription (85). The CarD^{WT}, CarD^{W85A}, and CarD^{K125A} proteins were all able to bind double stranded (dsDNA), as determined by a decrease in the unbound DNA band and the appearance of a slower migrating protein-bound DNA band in the EMSA (Fig. 1B). In

contrast, the CarD^{K90A}, CarD^{K90E}, and CarD^{K125E} mutants lost the ability to bind DNA, as determined by the dramatic decrease in the slower migrating protein-bound DNA band and the retention of the unbound DNA band in the EMSA (Fig. 1B). Therefore, the conserved lysine residues in the basic patch are important for binding DNA and single point mutations in the basic patch compromise this activity.

Disrupting the interaction between CarD and DNA does not affect the association of CarD with the RNAP but is detrimental for growth and viability. To investigate the role of the interaction between CarD and DNA *in vivo*, we attempted to replace the *carD* gene in *M. smegmatis* and *Mtb* with alleles encoding CarD^{K90A}, CarD^{K90E}, CarD^{K125A}, or CarD^{K125E} using a gene switching technique (101, 110). We successfully obtained all 4 mutants in *M. smegmatis* (Table 1). However, in *Mtb* we were only able to replace the *carD^{WT}* gene with the *carD^{K125A}* allele, which encodes the mutant that retains DNA binding activity *in vitro* during EMSA experiments (Fig. 1B and Table 1). In order to study the roles of the conserved tryptophan in CarD, we also attempted to replace the *carD* gene in *M. smegmatis* and *Mtb* with an allele encoding CarD^{W85A}. We were able to engineer *M. smegmatis* strains expressing the CarD^{W85A} mutant, but were unable to replace the *Mtb carD* allele with the *carD^{W85A}* allele (Table 1). We had previously shown that *Mtb* was more sensitive to interfering with the RNAP binding activity of CarD than *M. smegmatis* (101). These studies demonstrate that *Mtb* is also more sensitive to interfering with CarD's DNA binding activity and the conserved tryptophan, supporting the observation that *Mtb* has more stringent requirements for CarD's activities than *M. smegmatis*.

The possibility existed that changing the charge of the C-terminal basic patch would affect the stability or folding of CarD. In order to address this possibility, we constructed *M. smegmatis* strains expressing HA tagged CarD^{WT} or CarD mutants and analyzed the amount of

each mutant protein present in the cell by western blot (Fig. 2A) as a readout for protein stability. Using these same strains, we also performed co-immunoprecipitation experiments to determine if the CarD C-terminal point mutants retain the ability to associate with the RNAP β subunit (Fig. 2A) as a readout for proper folding. All of the C-terminal mutants were stably expressed in the bacteria and associated with RNAP β to the same degree as WT CarD except for $CarD^{K90E}$, which was barely detectable in the cell lysate suggesting that stability of the protein is affected by this substitution (Fig. 2A). Therefore, we chose not to pursue this mutant further. Alanine substitution at the conserved tryptophan ($CarD^{W85A}$) did not affect the amount of RNAP associated with CarD, demonstrating that not only does this residue not have a strong impact on binding dsDNA (Fig. 1B), it is also not involved in binding the RNAP (Fig. 2A) and indicates that this residue serves a unique function for CarD activity.

To determine whether the CarD-DNA interaction and the conserved tryptophan in CarD are important for optimal growth, we measured the growth rate of the viable *M. smegmatis* (Fig. 2B) and *Mtb* (Fig. 2C) CarD mutants. All strains used in this chapter contain only one *carD* allele. Based on at least 3 replicate experiments, *M. smegmatis* strains expressing the $CarD^{W85A}$, $CarD^{K90A}$, or $CarD^{K125E}$ mutant grew 1.23 (standard deviation, 0.084), 1.26 (standard deviation, 0.069), or 1.31 (standard deviation, 0.404) times slower than control strains expressing $CarD^{WT}$, respectively (Fig. 2B). This degree of decrease in growth rate is similar to what was observed previously with the $CarD^{R25E}$ mutant with weakened affinity to RNAP β (Fig. 2B and (101)). Therefore, even though these *M. smegmatis* strains are viable, efficient DNA binding and the conserved tryptophan are still required for optimal growth.

The K125A substitution in CarD, which did not affect DNA binding in EMSA experiments, slowed the growth of both *M. smegmatis* (Fig. 2B) and *Mtb* (Fig. 2C). Specifically,

the *M. smegmatis* CarD^{K125A} expressing strain grew 1.18 times slower than the control strain (standard deviation, 0.048) and the *Mtb* CarD^{K125A} expressing strain grew 1.35 times (standard deviation, 0.026) slower than the control strain. This indicates that this mutation affects CarD's function, despite maintaining WT levels of DNA binding in the EMSA assay.

The growth defects were even more apparent on solid media where the colonies of the *M. smegmatis* CarD^{R25E}, CarD^{W85A}, CarD^{K90A}, and CarD^{K125E} mutant expressing strains were smaller than those of the WT strain, as illustrated in Fig. 2D at the 10^{-3} and 10^{-4} dilutions. In addition, we noticed that the colonies from the strains expressing the CarD^{R25E} RID mutant and the CarD^{W85A} mutant had defects in pigmentation, while this was not observed in the strains expressing CarD mutants that loss the ability to bind dsDNA. This illustrates an important distinction between the conserved tryptophan and the lysine residues involved in interacting with DNA in that mutating the tryptophan residue does not phenocopy the DNA binding mutants but rather appears more like a RID mutant without affecting the interaction with RNAP. Therefore, this supports that the tryptophan and lysine residues within the C-terminus of CarD confer different functions for CarD.

Our discovery of CarD's DNA binding activity also raised the question of whether the previously described mutations in the RID domain that weakened the interaction between CarD and the RNAP β subunit (101) also affected the interaction with DNA. To examine this possibility, we investigated the DNA binding activity of the CarD^{R25E} and CarD^{R47E} mutants in EMSAs (Fig. 2E). These experiments showed that the CarD^{R25E} and CarD^{R47E} mutants retained the ability to bind and shift the *rrnA*PL DNA fragment, as evidenced by both the accumulation of a slower migrating complex containing the labeled DNA and the disappearance of the band corresponding to the unbound DNA fraction in these samples. The altered migration of the

CarD-DNA complexes formed by these mutants as compared to $CarD^{WT}$ may be due to the altered net charge in the CarD mutants or changes in cooperativity of the sequence nonspecific binding of CarD. Regardless, the retained DNA binding by these mutants indicates that these substitutions in the CarD RID do not affect the ability of CarD to bind DNA in the absence of RNAP. Therefore, the RNAP and DNA binding activities of CarD are able to function independently of each other.

The CarD-DNA interaction is dispensable for resistance to killing by oxidative stress. We have previously shown that both depletion of CarD and mutations that weaken the interaction between CarD and the RNAP dramatically compromise survival upon exposure to oxidative stress (Fig. 3, Chapter 1 Fig. 1, and (85, 101)), a stress encountered by *Mtb* during infection (115). To test whether the interaction between CarD and DNA is necessary for survival during oxidative stress, log-phase cultures of *M. smegmatis* strains expressing different *carD* alleles were treated with 25 mM H_2O_2 for 1 hour before dilutions were plated to count the surviving colony forming units (CFU) (Fig. 3). Strains expressing the $CarD^{K90A}$ or $CarD^{K125E}$ mutants, which lose the ability to bind DNA in EMSA experiments, were as resistant to oxidative stress as $CarD^{WT}$ expressing strains, indicating that the interaction between CarD and DNA is not required for the response to reactive oxygen species. These results further support that the CarD-RNAP and CarD-DNA interactions are independent, as mutations that weaken the interaction with RNAP but not with DNA cause sensitivity to reactive oxygen species. Interestingly, the W85A mutation in CarD sensitizes *M. smegmatis* to oxidative stress to a similar degree as mutations that weaken the interaction with RNAP (Fig. 3), but without affecting association with the RNAP (Fig. 2A). Therefore, the W85A substitution in CarD diverges from the CarD DNA binding mutants in the C-terminal basic patch in terms of the effect

on pigmentation and withstanding exposure to reactive oxygen species and we conclude that this conserved tryptophan is performing another function in CarD distinct from the RNAP and dsDNA binding activity of this protein. These data demonstrate that CarD encodes three activities that are distinct and operate independently of one another: 1) interaction with the RNAP, 2) interaction with dsDNA, and 3) an activity conferred by the conserve tryptophan at position 85. The role of CarD in combating oxidative stress requires a robust interaction with the RNAP β subunit and the conserved tryptophan residue, but not efficient binding to DNA.

A mutation in the DNA binding domain of CarD affects the pathogenesis of Mtb in murine tissues. We have shown that CarD is required for acute and chronic *Mtb* infection in mice (85) and that weakening the interaction between CarD and the RNAP results in a decrease in bacterial burden in both the lungs and the spleen during chronic infection (Chapter 1 Fig. 1, (101)). To investigate the effect of mutations in the DNA binding domain of CarD on pathogenesis, we infected C57BL/6 mice with *Mtb* strains expressing either $CarD^{WT}$ or the $CarD^{K125A}$ mutant. Although the K125A substitution did not affect DNA binding *in vitro*, since it is our only viable *Mtb* CarD C-terminal mutant, we used it to study the effect of mutations in the basic patch *in vivo*. The $CarD^{K125A}$ mutant displayed WT kinetics and levels of virulence during early acute infection in the lungs, but peaked at a bacterial burden half a log lower than the $CarD^{WT}$ expressing strain (Fig. 4A). The lower bacterial burden in animals infected with the *Mtb* $CarD^{K125A}$ mutant continued into the chronic phase of infection. The virulence defect of the $CarD^{K125A}$ strain during the chronic phase of infection was also apparent in the spleen where the titers of the $CarD^{K125A}$ expressing strain were more than a half log lower than the $CarD^{WT}$ expressing strain by 5 weeks post infection (Fig. 4B). The K125A substitution itself did not affect CarD binding to the DNA template tested *in vitro* (Fig. 1B). However, the loss of DNA

binding by the CarD^{K125E} mutant indicates that this lysine residue is involved in the association with DNA (Fig. 1B) and these data show that this lysine residue is also necessary for maintaining WT titers during chronic infection. Since both the CarD-RNAP and CarD-DNA interactions are required for maintaining high titers during chronic infection of mice (Fig. 4 and (101)), this excludes the source of attenuation being sensitivity to oxidative stress, since interfering with the ability of CarD to interact with DNA does not sensitize mycobacteria to reactive oxygen species (Fig. 3). Accordingly, we have previously shown that a CarD mutant with decreased affinity for the RNAP retains a virulence defect regardless of whether the mouse is able to mount a functional phagocytic oxidative burst (101).

Each of CarD's functional domains is important for resistance to clinically relevant antibiotics. We have previously shown that depleting CarD sensitizes *M. smegmatis* to killing by ciprofloxacin (85) and weakening the CarD-RNAP interaction increases the sensitivity of *M. smegmatis* to rifampicin and streptomycin (Chapter1 Fig. 1, (101)). To determine if the interaction between CarD and DNA is also necessary for antibiotic tolerance, we performed zone of inhibition assays and determined the sensitivity of *M. smegmatis* and *Mtb* strains expressing WT or mutant alleles of *carD* to clinically relevant antibiotics. We specifically tested sensitivity to ciprofloxacin (inhibition of DNA replication), rifampicin (inhibition of transcription), and streptomycin (inhibition of translation) since perturbations in CarD function have been associated with sensitivity to these drugs (85, 101). Each antibiotic targets a different essential cellular process, thereby testing the role of CarD's DNA binding activity in response to antibiotics with effects on diverse cellular processes. We found that mutations in the basic patch important for interacting with DNA (K90A and K125E) caused increased sensitivity to ciprofloxacin, rifampicin, and streptomycin in *M. smegmatis* in zone of inhibition assays (Fig.

5A-C), indicating that the association of CarD with DNA is important in tolerating treatment with these antibiotics. Both the K90A and the K125E mutations abolish the interaction between CarD and the *rrnA*PL DNA fragment in EMSAs (Fig. 1B), but expression of $CarD^{K125E}$ causes strains to be more sensitive to rifampicin than expression of $CarD^{K90A}$. This could be due to the more dramatic change in charge in the C-terminus of CarD in the K125E mutant. Expression of the $CarD^{K125A}$ mutant, which retains binding to the DNA template tested *in vitro* (Fig. 1B), also increases sensitivity of *M. smegmatis* to the antibiotics tested. The effect of the K125A mutation in CarD on antibiotic sensitivity, growth rate, and *Mtb* virulence indicates that this lysine residue is important for an activity that is affected by the alanine substitution. Using similar assays, we observed that the conserved tryptophan residue at position 85 in CarD is also important for tolerance to rifampicin, ciprofloxacin, and streptomycin in *M. smegmatis* and the W85A mutation results in sensitivity to these antibiotics that is comparable to mutations that either weaken the interaction with DNA (Fig. 5A-C).

Previous experiments that examined the antibiotic sensitivity of *M. smegmatis* strains expressing CarD mutants with weakened affinity to the RNAP employed transient treatment assays in liquid cultures (85, 101). Since we were now using a zone of inhibition assay where the bacteria are continuously exposed to the antibiotics from a disk on solid agar, we also confirmed that the mutations that weaken the interaction between CarD and the RNAP (R25E and R47E) also increase the sensitivity of *M. smegmatis* to ciprofloxacin (Fig. 5A), rifampicin (Fig. 5B), and streptomycin (Fig. 5C) in these assays. These data demonstrate that the role of CarD during tolerance to these antibiotics requires all three of CarD's proposed activities. This is distinct from the response to oxidative stress, where CarD's ability to bind DNA is dispensable.

We had previously reported that depletion of CarD but not mutations in the RID domain increased sensitivity to ciprofloxacin in transient liquid kill assays in which exponential cultures of *M. smegmatis* were treated for 2 hours with antibiotic and then survival was monitored by plating the surviving CFUs (85, 101). Despite the increased sensitivity of all of the CarD mutants to ciprofloxacin in the zone of inhibition assays, we found no significant differences in sensitivity after 2 hours of antibiotic treatment in liquid culture (Fig. 5D), which is similar to previous reports. The difference between these results may lie in the experimental design where bacteria are only transiently exposed to the antibiotic in liquid cultures as opposed to continuous exposure in the zone of inhibition assays. These data suggest that, unlike cells depleted for CarD, strains encoding mutant alleles of *carD* are able to mount enough of a response to survive transient exposure to ciprofloxacin, but this response is less effective than in bacteria expressing $CarD^{WT}$, thus resulting in sensitivity to prolonged exposure.

Each of CarD's functional domains is important for regulation of rRNA levels. We have shown that CarD is localized to promoters throughout the genome, indicating that it is likely a global transcriptional regulator (Chapter 1 Fig. 2, (87)). During initial characterization of CarD, a microarray was performed that compared the transcriptome of *M. smegmatis* expressing CarD to that of *M. smegmatis* depleted of CarD protein. The most striking change in the transcriptome upon depletion of CarD was observed in transcripts related to the translational machinery, including rRNA itself (85). We have chosen to first characterize CarD's effect on regulation of rRNA promoters due to the exquisite sensitivity of the translation machinery to CarD depletion and the physiological importance of rRNA regulation. We predicted that the CarD mutants with weakened affinity to DNA or RNAP or with mutation of the conserved tryptophan display slower growth kinetics and increased sensitivity to antibiotics due to improper

transcriptional regulation by CarD, including transcriptional regulation of the translational machinery. To determine if the weakening CarD's interactions with RNAP or DNA or mutation of the conserved tryptophan affect CarD's regulation of rRNA, we performed quantitative real-time PCR (qRT-PCR) to measure the levels of 16S rRNA in exponentially growing cultures of *M. smegmatis* and *Mtb* expressing WT or mutant alleles of *carD*. In *M. smegmatis*, mutations in CarD that affected the interaction with RNAP, the association with DNA, or the conserved tryptophan each led to decreased levels of rRNA in the bacteria, indicating that all of CarD's activities are involved in regulating rRNA transcription (Fig. 6A). The strain expressing $CarD^{K125E}$ had a more severe defect in 16S rRNA transcript levels than the $CarD^{K90A}$ expressing strain, which mirrors the trend in survival defects during exposure to rifampicin (Fig. 5B). The importance of the lysine residue at position 125 was further confirmed when the $CarD^{K125A}$ mutant also had decreased levels of rRNA in *M. smegmatis* and *Mtb* as compared to $CarD^{WT}$ (Fig. 6A-B). Decreasing CarD's affinity for RNAP also decreased 16S rRNA levels in *Mtb* (Fig. 6B).

CarD activates transcription initiation at rRNA promoters and requires each of its functional domains for this activity. CarD has been shown to localize to promoters throughout the genome (Chapter 1 Fig. 2, (87)) which supports a role for CarD in regulating transcription initiation, however this hypothesis has not been tested. To directly test the effect of CarD on rRNA promoter activity *in vivo*, we measured β-galactosidase (βgal) activity from *lacZ* fused to the rRNA promoters. In *Mtb* there is only one rRNA operon, *rrnA*, and transcription is driven by two promoters, P1 and P3 (116, 117). Transcription of the homologous *rrnA* operon in *M. smegmatis* is derived from similar P1 and P3 promoters as well as an additional P2 promoter (116, 117). Individual promoters from the *M. smegmatis rrnA* operon were cloned upstream of

the *lacZ* gene, transformed into *M. smegmatis* strains expressing WT or mutant alleles of *carD*, and β-galactosidase activity resulting from the expression of *lacZ* was measured to determine promoter activity in each *M. smegmatis* strain (Fig. 6C). Interfering with the ability of CarD to bind RNAP or DNA as well as mutation of the conserved tryptophan decreased transcription from *M. smegmatis rrnA* promoters P1 and P2 (Fig. 6C). This result indicates that $CarD^{WT}$ activates transcription from these promoters using a mechanism that is dependent on each of its functional domains. β-galactosidase activity measured from the *rrnA*P3 promoter was highly variable between replicates and any differences were not statistically significant, therefore we could not make conclusions regarding the role of CarD's activities at this promoter in the *lacZ* fusion experiments. The mutations in the CarD DNA binding domain (K90A, K125A, and K125E) had more dramatic effects on transcription from the *rrnA*P2 promoter as compared to the *rrnA*P1 promoter, resulting in a greater decrease in β-galactosidase activity from that seen in $CarD^{WT}$ strains at the *rrnA*P2 promoters (Fig. 6C). This difference suggests that the strength of interaction between CarD and DNA is more important for initiation of transcription from the *rrnA*P2 promoter than from the *rrnA*P1 promoter. Conversely, the R25E and the W85A mutations in CarD lead to a similar fold decrease at both promoters (0.47 and 0.41 fold lower transcription in $CarD^{R25E}$ expressing strains and 0.31 and 0.35 fold lower in $CarD^{W85A}$ expressing strains for *rrnA*P1 and *rrnA*P2, respectively; Fig. 6C), indicating that the promoters require the same levels of the CarD-RNAP interaction and the activity conferred by the conserved tryptophan. The $CarD^{K125E}$ mutant also had a more dramatic defect in activating both the *rrnA*P1 and *rrnA*P2 promoter as compared to the $CarD^{K90A}$ mutant, which agrees with the observation that the levels of 16S rRNA were lower in the $CarD^{K125E}$ expressing strain as compared to the $CarD^{K90A}$ expressing strain (Fig. 6A and C). These data demonstrate that each of CarD's three

activities are critical for the activation of rRNA promoters by CarD. Unlike the other functions encoded by CarD, its interaction with DNA is the first activity ascribed to CarD that may exhibit some promoter specificity in terms of the importance of its activity.

CarD requires its interactions with RNAP and DNA as well as the activity of the conserved tryptophan to stabilize RNAP-promoter complexes. The lower levels of rRNA and *rrnA* promoter activity (Fig. 6A-C) in mycobacterial strains expressing CarD mutants indicates both that CarD is regulating transcription initiation and that these activities are important for mechanism by which CarD regulates transcription. To investigate the mechanism of transcriptional regulation by CarD, we performed single round *in vitro* transcription assays. Specifically, mycobacterial RNAP σ^A holoenzyme was incubated in the presence or absence of CarD with circular supercoiled DNA plasmids containing the *Mtb rrnA*P3 promoter driving transcription of a test transcript to allow for formation of transcription initiation complexes. dsDNA competitor was added to compete away free RNAP and prevent formation of new RNAP-promoter complexes. NTPs were added at successive time points after the addition of competitor and the amount of transcript produced was measured. As transcripts can only be initiated from transcription competent complexes formed prior to the addition of competitor, by monitoring the amount of transcription that occurs at different times following the addition of competitor, the half-life of the RNAP-promoter complexes can be calculated. Addition of $CarD^{WT}$ to single round *in vitro* transcription assays increased the half-life of the RNAP-promoter complexes formed at the *Mtb rrnA*P3 promoter by 14.7 fold (Fig. 7A-C) indicating that $CarD^{WT}$ stabilizes a competitor resistant RNAP-promoter complex formed during transcription initiation.

In the above experiment, the *Mtb rrnA*P3 promoter was present in isolation of its natural genomic context. In the genome, the beginning of the -35 region of *rrnA*P3 is located 48 bp downstream from the end of the -10 region of *rrnA*P1. It is possible that the *in vitro* transcription assays with only one promoter neglect effects of the adjacent *rrnA*P1 promoter on the stability of initiation complexes formed at *rrnA*P3, which may alter how the initiation complexes respond to CarD activity. To study the stability of RNAP-promoter complexes formed at the *Mtb rrnA*P3 promoter in its native context, we generated DNA templates containing nucleotides 1,469,982-1,470,234 of the *Mtb* genome, which harbored both of the *rrnA* promoters and their intervening sequences. We designated this DNA fragment *Mtb rrnA*P13. *rrnA*P1 and *rrnA*P3 each produce a transcript of a unique length that can be quantified independently (Fig. 7A). However, the *rrnA*P1 promoter is relatively weak (116, 117) and transcription from the *rrnA*P1 promoter is not reliably detected in our assays. Single round *in vitro* transcription assays showed that in the absence of CarD, the RNAP-promoter complexes formed at *Mtb rrnA*P3 were 3 fold more stable when *rrnA*P1 was located upstream (Fig. 7C). Therefore, the upstream *rrnA*P1 promoter has a polar effect on transcription initiation complexes that form at *rrnA*P3. The addition of $CarD^{WT}$ to the reactions containing the *rrnA*P13 construct further stabilized the half-life of the RNAP-promoter complexes formed at *rrnA*P3 (Fig. 7C). Interestingly, the addition of $CarD^{WT}$ resulted in a similar final RNAP-promoter complex half-life regardless of the context of *rrnA*P3. This data indicates that CarD has a maximum effect on the stability of the RNAP-promoter complexes at *rrnA*P3 that is independent of the stabilizing effect of the upstream *rrnA*P1 promoter (Fig. 7C).

Using our point mutant collection, we investigated the contribution of each of CarD's activities to its ability to stabilize RNAP-promoter complexes at the *Mtb rrnA*P3 promoter in the context of the *rrnA*P13 construct (Fig. 7D). Addition of $CarD^{WT}$ increased the half-life of the

RNAP-promoter complex at the *Mtb rrnA*P3 promoter 5.3 fold relative to the half-life of the complex in the absence of CarD (Fig. 7D). The residues involved in the interaction between CarD and the RNAP as well as the W85 residue in CarD were essential for the ability of CarD to stabilize RNAP-promoter complexes at *Mtb rrnA*P3 in the *rrnA*P13 construct, and mutations in these residues abolish this activity at this promoter (Fig. 7D). CarD mutants with decreased affinity for DNA, CarD^{K90A} and CarD^{K125E}, were also attenuated in their ability to stabilize the RNAP-promoters complexes at *rrnA*P3 (Fig. 7D). These data demonstrate that each of CarD's functional domains are required for CarD's ability to stabilize RNAP-*rrnA*P3 complexes, which could explain the lower levels of rRNA in strains expressing CarD mutants (Fig. 6A).

Similar experiments were also performed to assay the effect of CarD on the stability RNAP-promoter complexes formed at *M. smegmatis rrnA*P3 in the context of the *rrnA*P123 construct, which harbors nucleotides 5,029,667-5,029,905 of the *M. smegmatis* genome thus encoding all 3 *rrnA* promoters (Fig. 7A). Unlike CarD^{WT}, the CarD^{R25E}, CarD^{K90A}, and CarD^{K125E} mutants were unable to increase the half-life of RNAP-promoter complexes at *M. smegmatis rrnA*P3, demonstrating that the CarD-RNAP and CarD-DNA interactions are required for CarD's activity at *M. smegmatis rrnA*P3 (Fig. 7E). The CarD^{W85A} mutant had only a partial defect in the ability to stabilize complexes at this promoter as compared to CarD^{WT} which may partially explain the essentiality of this mutation in *Mtb* but not *M. smegmatis*.

Discussion

In this chapter, we have dissected three independent activities encoded by CarD: binding the RNAP, binding dsDNA, and the yet to be elucidated function of the conserved tryptophan residue within the C-terminal basic patch. We have generated a panel of mutants that individually affect one of these activities and have used them to probe the functional importance

of each activity by both *in vitro* and *in vivo* assays. Biochemically, we have demonstrated that CarD stabilizes a competitor resistant RNAP-promoter complex formed during transcriptional initiation by a mechanism that is dependent upon all three of CarD's activities (Fig. 7D-E). Mutating any single activity of CarD also compromises CarD's ability to activate transcription of rRNA promoters *in vivo*, resulting in decreased transcription from rRNA promoters in transcriptional fusion assays (Fig. 6C) and lower rRNA levels (Fig. 6A-B). Furthermore, we have shown that all three of CarD's activities are required for optimal growth rate and resistance to diverse stresses (Figs. 2 and 5).

Much of the characterization of CarD's role in mycobacteria has been focused on its importance in regulating rRNA transcription. The inability of CarD mutants to activate transcription of rRNA promoters in transcriptional fusion experiments (Fig. 6C) as well as the lower levels of 16S rRNA in CarD mutants (Fig. 6A-B) indicates that CarD functions as a transcriptional activator of these promoters. In all bacteria, the rates of ribosomal protein transcription, ribosome biogenesis, and cell growth correlate to the levels of rRNA production (118–120). One aspect of *Mtb* biology that contributes to its success as a pathogen is its ability to persist within the host by entering a state of dormancy (121, 122), which most likely requires stringent regulation of ribosome biogenesis and translation. However, the regulation of rRNA transcription in mycobacteria and its relationship to pathogenesis are under-investigated. Investigations into the mechanism by which CarD regulates rRNA transcription will shed light onto these important topics.

An important advance in our understanding of the regulation of rRNA transcription by CarD was a kinetic analysis of CarD regulation of transcription at a rRNA promoter performed by our collaborators in the Galburt lab (109). In this study, a Cy3 fluorophore is placed at the +2

position relative to the transcription start site of the non-template strand of the *Mtb rrnA*P3 promoter. The fluorophore's fluorescence is enhanced 2-fold in RP_o as compared to the free DNA or in RP_c, allowing for monitoring of open complex formation and equilibrium in real time (109). This study and others have shown that, compared to the *E. coli* RNAP, the mycobacterial RNAP is deficient at formation of RP_o and that CarD enhances this activity such that the *M. bovis* RNAP more closely resembles the *E. coli* RNAP (109, 123). The fluorescence assay was the first study to determine the kinetics of CarD's effect on open complex formation. This kinetic analysis combined with the *in vitro* and *in vivo* dissection of CarD activity in this paper has led to the following model of CarD activity (Fig. 8A). With high affinity, CarD interacts with RP_o where its interaction with RNAP and DNA as well as the activity of the conserved tryptophan stabilizes the complex by reducing the reverse rate of collapse back to closed complex. With lower affinity, CarD interacts with RP_c and increases the forward rate of promoter opening. Analysis of CarD mutants with decreased affinity for RNAP or DNA or with mutation of the conserved tryptophan were all less able to stabilize the open complex in this assay. Importantly, I determined the cellular concentration of RNAP β, RNAP σ, and CarD in *Mtb* and found that the concentrations within the cell are similar to concentrations used in the fluorescence assay suggesting that CarD both inhibits promoter collapse and enhances opening *in vivo* (Fig. 8B, (109)).Our studies of single point mutants of CarD illustrate the importance of CarD-mediated regulation on establishing a gene expression profile that contributes to growth and stress resistance.

While current data supports a role for CarD as an activator of rRNA transcription, investigations into CarD activity at non-ribosomal promoters has yet to be performed. As CarD is localized to promoters throughout the genome (87), it remains possible that CarD may not

always function as a transcriptional activator. For instance, if an RNAP holoenzyme has a high affinity for a promoter in the absence of CarD, further increasing the stability of this complex could repress transcription by inhibiting promoter escape by RNAP. The ability of a transcription factor to activate or repress transcription in this manner has been demonstrated for the phage ϕ29 protein p4 which represses strong promoters but activates weak promoters by binding upstream of the promoter and interacting with the C-terminal domain of the RNAP α subunit (124).

CarD homologs are essential in *Mtb*, *M. sme*gmatis (85) and *Myxococcus xanthus* (125) and knockouts were not attainable in *Borrelia burgdorferi* (126). However, the reasons for CarD essentiality in these bacteria and how CarD confers antibiotic and stress tolerance in mycobacteria remain unknown. It is possible that CarD is essential due to its role in regulating rRNA transcription, however more studies are necessary to determine if this is the case. Since CarD is associated with all RNAP-σ^A transcription initiation complexes within the *M. smegmatis* genome (87), it is likely that CarD is a general component of the transcription initiation machinery rather than regulation of specific transcripts. Organisms lacking CarD, such as *E. coli*, must have an alternative means to fulfill CarD's role in transcription initiation. This may be accomplished if the bacteria lacking CarD encode a different accessory protein that renders CarD unnecessary or if there are inherent differences in the transcription complexes of these bacteria that allow them to function efficiently in the absence of CarD. Recent data on the inherent differences between the *E. coli* RNAP, which forms stable RP_o's, and the RNAPs of *M. bovis*, *Thermus thermophilus*, *Thermus aquaticus*, and *Bacillus subtilis* which all form unstable RP_o's suggest that there is more variation in bacterial RNAP than previously understood (109, 123, 127–129). Interestingly, there are CarD homologs in both *B. subtilis* and *Thermus spp*. suggesting unstable RP_o formation may be linked the maintaining a *carD* homolog (85–87).

However, *carD* is not essential in *B. subtilis* (personal communication with Dr. Jonathan Dworkin), suggesting the stability of RP_o is not the only determinant of CarD essentiality (130). Our data has shown that all three functional modules of CarD are essential in *Mtb* but not in *M. smegmatis*. This implies that there is a more stringent requirement for CarD's function in *Mtb* as compared to *M. smegmatis*. As an obligate pathogen, *Mtb* has evolved for its very specialized niche, whereas *M. smegmatis* and other environmental mycobacteria must be more versatile and generally contain larger genomes with greater redundancy in stress response pathways. Therefore, it is possible that *M. smegmatis* expresses other factors that allow for loss of the functions conferred by the conserved tryptophan and the DNA binding activity in CarD. Further comparison of the physiology of these two mycobacterial species could provide insight into the reasons behind the essentiality of CarD.

Of the CarD mutants with amino acid substitutions in the C-terminus, we were only able to obtain a $CarD^{K125A}$ expressing strain in *Mtb*. The $CarD^{K125A}$ mutant retains WT levels of interaction with DNA in EMSA assays (Fig. 1B) and WT levels of association with the RNAP in co-immunoprecipitation assays (Fig. 2A) but results in decreased growth (Fig. 2B-C), decreased rRNA levels (Fig. 6A-B), and increased sensitivity to antibiotic treatment (Fig. 5A-C). Therefore, the K125A mutation does have functional consequences for CarD. These consequences could be the result of this mutation affecting the binding to sequences other than what was used in the EMSA or a defect in DNA binding that is undetectable by the EMSA. It also remains possible that the mutation affects an unknown function of CarD distinct from its interaction with either RNAP or DNA.

One of the most surprising findings from our studies was that while the majority of CarD mutant phenotypes are similar regardless of which of CarD's activities were targeted, increased

sensitivity to H_2O_2 and loss of pigmentation are specific for mutations at the conserved tryptophan or that weaken the interaction with RNAP (Fig. 3). This suggests that either dsDNA binding is dispensable for regulation of transcripts required for tolerating H_2O_2 treatment or CarD has a cellular function independent of its interaction with DNA but necessary to survive oxidative stress. Because the mutation of the conserved tryptophan results in sensitivity to reactive oxygen species without affecting the interaction with RNAP (Fig. 2A), it is unlikely that the CarD-mediated resistance is due to CarD physically shielding the RNAP from oxidative damage. The differential requirement for the association with the RNAP but not the DNA to withstand oxidative stress could be linked to the observation that the CarD/DNA interaction, but not the CarD/RNAP interaction, exhibits some promoter specificity in terms of the importance of its activity. However, the mechanism by which H_2O_2 is killing mycobacteria is currently unknown and the reason for increased sensitivity in some CarD mutants remains elusive.

Acknowledgements

We thank Dr. Bob Landick for assistance with protocols and reagents related to the isolation of *M. bovis* RNAP-σ^A holoenzyme and Drs. Seth Darst and Elizabeth Campbell for structural studies that help identify residues involved in DNA binding by CarD. C.L.S. is supported by a Biomedical Research Grant from the American Lung Association and C.L.S. and E.A.G are supported by Grant GM107544 from the National Institutes of Health. A.L.G. is supported by the NIGMS Cell and Molecular Biology Training Grant GM007067 and the Stephen I. Morse Graduate Fellowship.

Table 1: CarD mutants studied in this chapter

Mutation in CarD	***M. smegmatis***	***M. tuberculosis***	**Activity Affected by Mutation**
R25E	Viable	Unattainable	Interaction with RNAP
R47E	Viable	Viable	Interaction with RNAP
W85A	Viable	Unattainable	Unknown
K90A	Viable	Unattainable	Interaction with dsDNA
K90E	Viable	Unattainable	Interaction with dsDNA
K125A	Viable	Viable	None Determined
K125E	Viable	Unattainable	Interaction with dsDNA

Figure Legends

Figure 1. CarD's C-terminal basic patch is responsible for its interaction with DNA.

A. Electrostatic surface diagram of *Mtb* CarD modified from (87) illustrating the location of W85, K90, and K125 within the basic patch.

B. Image of a nondenaturing polyacrylamide gel from an EMSA with no protein, *M. smegmatis* $CarD^{WT}$, $CarD^{W85A}$, $CarD^{K90A}$, $CarD^{K90E}$, $CarD^{K125A}$, or $CarD^{K125E}$ incubated with IRDye labeled *M. smegmatis rrnA*PL DNA. The reactions were separated on a nondenaturing polyacrylamide gel, which was then imaged using the Odyssey CLX imaging system (LI-COR).

Figure 2. Each of CarD's functional domains is required for optimal growth in mycobacteria.

A. Immunoprecipitation experiments with a monoclonal antibody specific for HA in the *M. smegmatis* strains expressing $CarD^{WT}$-HA (lane 1), $CarD^{R25E}$-HA (lane 2), $CarD^{K90A}$-HA (lane 3), $CarD^{K90E}$-HA (lane 4), $CarD^{W85A}$-HA (lane 5), $CarD^{K125A}$-HA (lane 6), or $CarD^{K125E}$-HA (lane 7). Inputs (before immunoprecipitation) and eluates were analyzed by western blotting with antibodies specific for either RNAP β (Panel A and D) or CarD (Panel B, C, and E). Panel C is a longer exposure of the film from panel B in order to show the $CarD^{K90E}$ band.

B-C. The doubling time of each CarD strain was expressed as a ratio to the doubling time of the $CarD^{WT}$ expressing strain for the *M. smegmatis* strains expressing $CarD^{WT}$, $CarD^{R25E}$, $CarD^{K90A}$, $CarD^{W85A}$, $CarD^{K125A}$, or $CarD^{K125E}$ (B) and *Mtb* strains expressing $CarD^{WT}$, $CarD^{K125A}$, or $CarD^{R47E}$ (C). Each graph shows the mean ± SEM of data from at least three replicates. Significance of the differences between mutant

strains and WT were determined by calculating *P* values by Student's *t* test. An asterisk indicates significance with a *P* value of <0.05, two asterisks indicate significance with a *P* value of <0.01, and three asterisks indicate significance with a *P* value of <0.005.

D. Plated dilutions of *M. smegmatis* strains expressing $CarD^{WT}$, $CarD^{R25E}$, $CarD^{W85A}$, $CarD^{K90A}$, $CarD^{K125A}$, or $CarD^{K125E}$ on LB after 3 days of growth at 37°C.

E. Image of nondenaturing polyacrylamide gel from EMSAs with no protein, *Mtb* $CarD^{WT}$, $CarD^{R25E}$, or $CarD^{R47E}$ incubated with IRDye labeled *M. smegmatis rrnA*PL DNA. The reactions were separated on a nondenaturing polyacrylamide gel, which was then imaged using the Odyssey CLX imaging system (LI-COR).

Figure 3. The CarD-DNA interaction is dispensable for resistance to killing by oxidative stress. Log-phase *M. smegmatis* strains expressing $CarD^{WT}$, $CarD^{R25E}$, $CarD^{R47E}$, $CarD^{W85A}$, $CarD^{K90A}$, $CarD^{K125A}$, or $CarD^{K125E}$ in LB were treated for 1 hr with 25 mM H_2O_2. After treatment, dilutions were plated on LB and the surviving CFUs were counted. Survival of each replicate is graphed as a ratio of CFU in treated cultures to that in untreated cultures along with the mean ± SEM of the ratios for each set of replicates. Each sample is represented by a black circle.

Figure 4. A mutation in the DNA binding domain of CarD affects the pathogenesis of *Mtb* in murine tissues. C57BL/6 mice were infected by the aerosol route with the *Mtb* strains expressing either $CarD^{WT}$ (black circles) or $CarD^{K125A}$ (grey squares). Shown are bacterial titers in the lungs (A) and spleens (B) of the infected mice. Each time point is the mean ± SEM of data from 6 mice per strain, combined from two experiments.

The significance of the differences between the CarD^{K125A} and the CarD^{WT} strain in both panels were determined as described for Figure 2.

Figure 5. Each of CarD's functional domains is important for resistance to clinically relevant antibiotics.

A-C. Survival of *M. smegmatis* strains during the disk zone of inhibition assays. Five hundred microliters of log-phase *M. smegmatis* strains expressing CarD^{WT}, CarD^{R25E}, CarD^{R47E}, CarD^{W85A}, CarD^{K90A}, CarD^{K125A}, or CarD^{K125E} were plated on LB agar, and a disk spotted with 5µl of 1 mg ml^{-1} ciprofloxacin (A), 100 mg ml^{-1} rifampicin (B), or 200 mg ml^{-1} streptomycin (C) was placed in the middle of the bacterial lawn. A representative experiment is shown above the x-axis. The radius of the zone of inhibition for each replicate and the mean ± SEM for each set of replicates is graphed, with each sample represented by a black circle.

D. Survival of *M. smegmatis* strains during transient ciprofloxacin treatment. Log-phase cultures of the *M. smegmatis* strains expressing CarD^{WT}, CarD^{R25E}, CarD^{R47E}, CarD^{W85A}, CarD^{K90A}, CarD^{K125A}, or CarD^{K125E} growing in LB were treated for 2 hr with 10 µg ml^{-1} ciprofloxacin before the dilutions were plated to determine the surviving CFU. Graphed is the ratio of CFUs in treated cultures to that in untreated cultures for each replicate and the mean ± SEM for each set of replicates, with each sample represented by a black circle.

For all panels, the significance of differences between mutant strains and WT were determined as described in Figure 2.

Figure 6. Each of CarD's functional domains are important for transcriptional regulation.

A-B. 16S rRNA levels in log phase cultures of *M. smegmatis* strains expressing CarD^{WT}, CarD^{R25E}, CarD^{R47E}, CarD^{W85A}, CarD^{K90A}, CarD^{K125A}, or CarD^{K125E} grown in LB (**A**) and *Mtb* strains expressing CarD^{WT} or CarD^{K125A} grown in 7H9 (**B**) as determined by qRT-PCR and expressed as a ratio to the levels in CarD^{WT} expressing strains. Each graph shows the mean ± SEM of data from at least 3 replicates and the significance of the differences between mutant strains and WT.

C. *M. smegmatis* strains expressing CarD^{WT}, CarD^{R25E}, CarD^{R47E}, CarD^{W85A}, CarD^{K90A}, CarD^{K125A}, or CarD^{K125E} were transformed with the promoter-*lacZ* fusion constructs diagramed above each graph. The ratio of β-galactosidase activity that resulted from *lacZ* expression from the *M. smegmatis rrnA*P1, *rrnA*P2, or *rrnA*P3 promoter in each fusion in the CarD mutant versus CarD^{WT} expressing strains was calculated and graphed as the mean of at least 3 replicates ± SEM. Shown above each bar is the significance of the difference between the mutant strain and WT from the same promoter. The bracket designates the significance of the difference between the activation of the *rrnA*P1 and *rrnA*P2 promoters in the CarD^{K125E} expressing strain.

For all panels, the significance of differences were calculated as described in Figure 2.

Figure 7. CarD's interaction with DNA is important for stabilization of RNAP-promoter complexes.

A. Schematic of the *Mtb* and *M. smegmatis* rRNA promoter constructs used during *in vitro* transcription assays. The length of the *rrnA*P3-derived transcripts initiated at the promoter and ending at the terminator (T) are shown.

B. Autoradiographs of ^{32}P-labeled transcripts on denaturing polyacrylamide gels from a representative complex stability assay using *M. bovis* RNAP-σ^A holoenzyme and the

*rrnA*P3 construct. The amount of transcript formed at the indicated time-points following the addition of competitor is shown when no factor is added versus when $CarD^{WT}$ is added. The intensity of each band was quantified and the ratio to the intensity at time 0 was calculated and annotated under the gel.

C. The half-life ($t_{1/2}$) of RNAP-promoter complexes formed at the *rrnA*P3 promoter in the *Mtb rrnA*P3 and the *Mtb rrnA*P13 constructs in the presence (hatched bars) and absence (solid grey bars) of $CarD^{WT}$ was calculated in GraphPad Prism based on the amount of transcripts formed during complex stability assays. The graph shows the mean of at least 3 replicates ± SEM and the significance of the differences in the comparisons designated by the brackets are shown.

D-E. Ratio of the $t_{1/2}$ of RNAP-promoter complexes containing $CarD^{WT}$, $CarD^{R25E}$, $CarD^{W85A}$, $CarD^{K90A}$, or $CarD^{K125E}$ to the $t_{1/2}$ of RNAP-promoter complexes in the absence of CarD as determined by single round complex stability *in vitro* transcription assays at the *rrnA*P3 promoter on the *Mtb rrnA*P13 construct (D) or the *M. smegmatis rrnA*P123 construct (E). The significance of differences as compared to the reactions containing no factor are shown.

For all panels, the significance of differences were determined as described for Figure 2.

Figure 8. Model of CarD activity at promoters.

A. With high affinity, CarD interacts with the promoter DNA (1) and RNAP β (2) of RP_o. These interactions stabilize RP_o, which decreases the reverse rate of collapse to RP_c (3). With lower affinity, CarD $_{interacts}$ with RP_c and promotes formation of RP_o (4). Together, these effects of CarD alter gene expression to improve viability and resistance to stress (5).

B. The cellular concentration of RNAP β, RNAP σ^A, and CarD in WT *M. tuberculosis*. Published in (109).

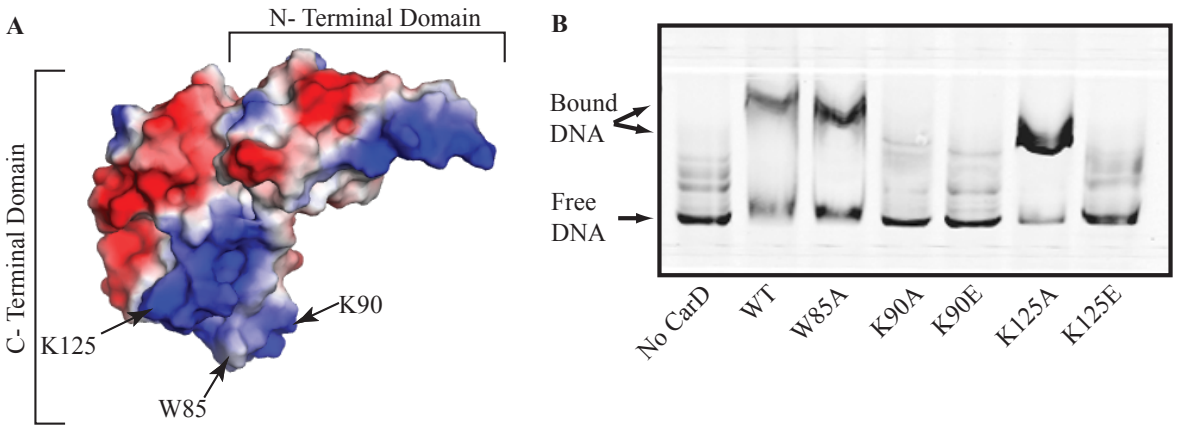

A
N- Terminal Domain
C- Terminal Domain
K125
K90
W85
B
Bound
DNA
Free
DNA
No CarD
WT
W85A
K90A
K90E
K125A
K125E

Figure 1

Figure 2

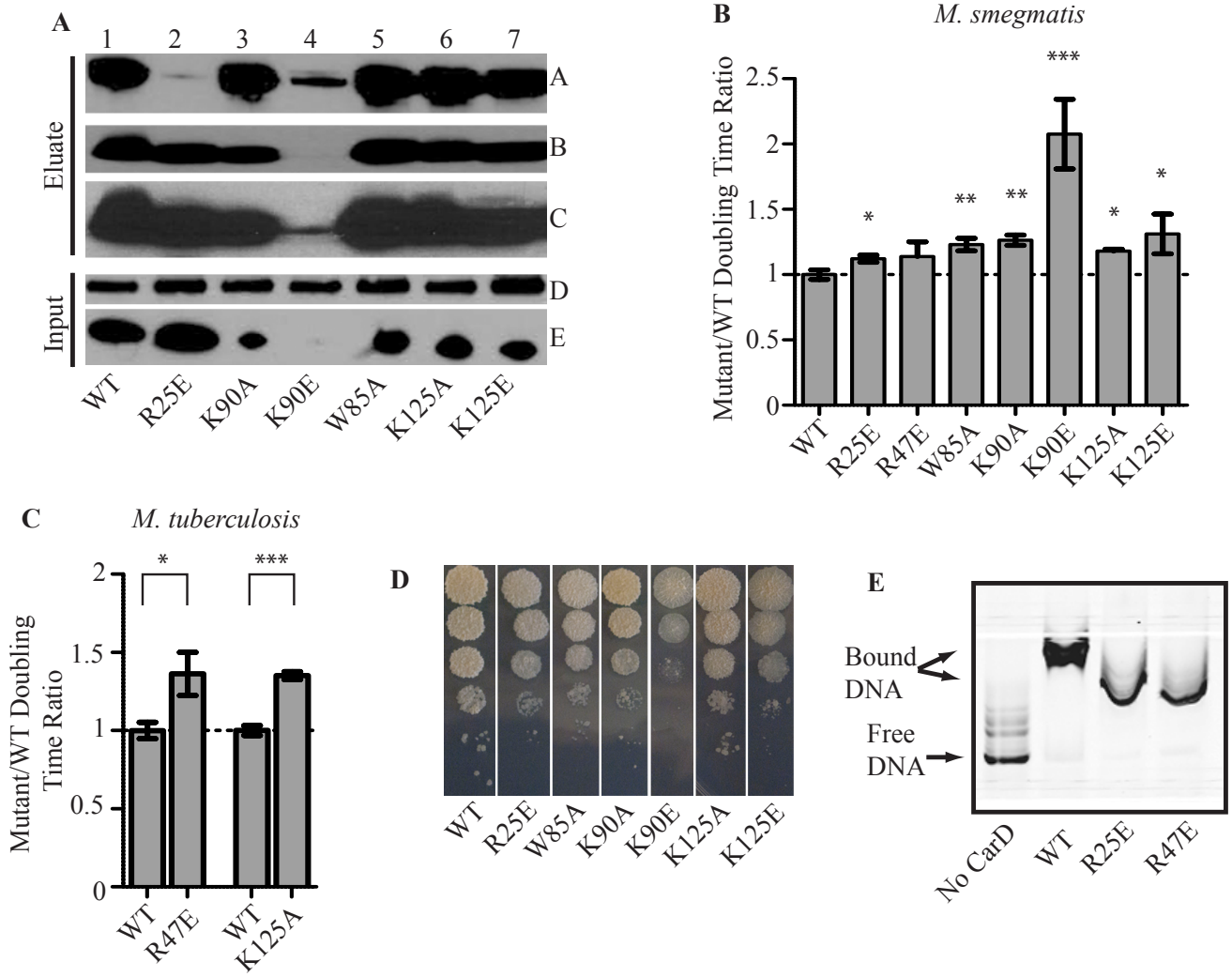

A
1 2 3 4 5 6 7
Eluate
Input
A
B
C
D
E
WT
R25E
K90A
K90E
W85A
K125A
K125E
B
M. smegmatis
Mutant/WT Doubling Time Ratio
2.5
2
1.5
1
0.5
0
*
**
**
***
*
*
WT
R25E
R47E
W85A
K90A
K90E
K125A
K125E
C
M. tuberculosis
Mutant/WT Doubling Time Ratio
2
1.5
1
0.5
0
*
***
WT
R47E
WT
K125A
D
WT
R25E
W85A
K90A
K90E
K125A
K125E
E
Bound DNA
Free DNA
No CarD
WT
R25E
R47E

Figure 3

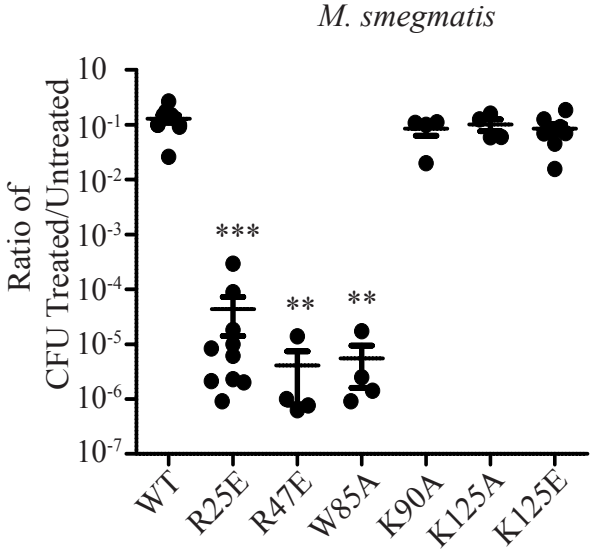

M. smegmatis
Ratio of CFU Treated/Untreated
10
10-1
10-2
10-3
10-4
10-5
10-6
10-7
***
**
**
WT
R25E
R47E
W85A
K90A
K125A
K125E

Figure 4

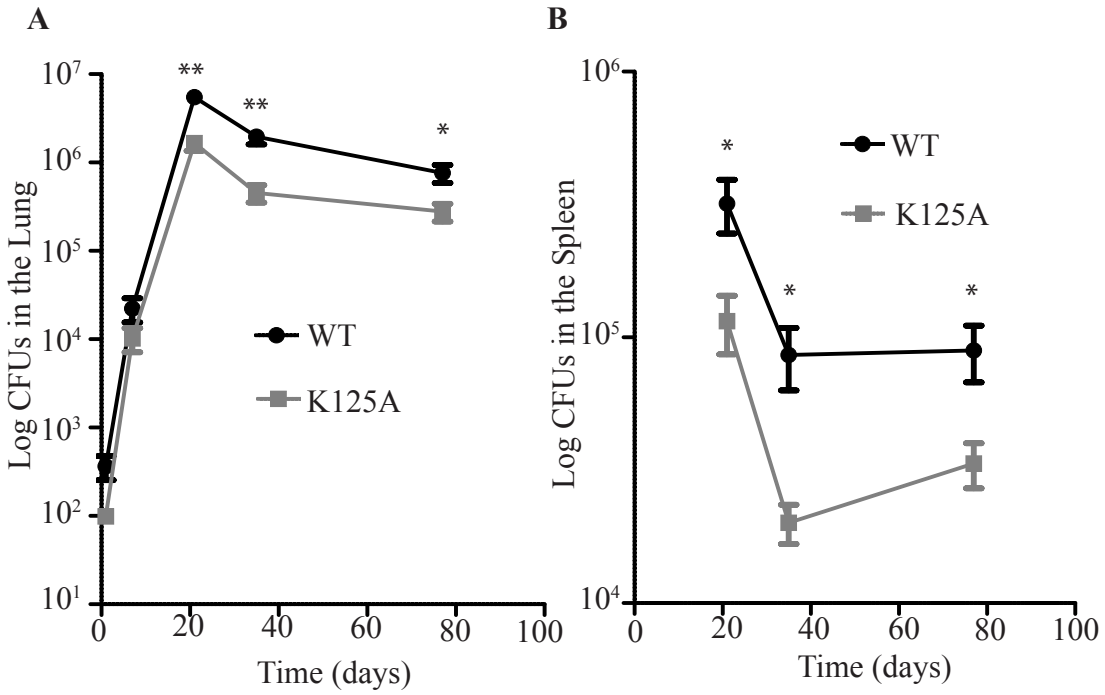

A
B
Log CFUs in the Lung
Log CFUs in the Spleen
Time (days)
Time (days)
WT
K125A
WT
K125A
**
**
*
*
*
*

Figure 5

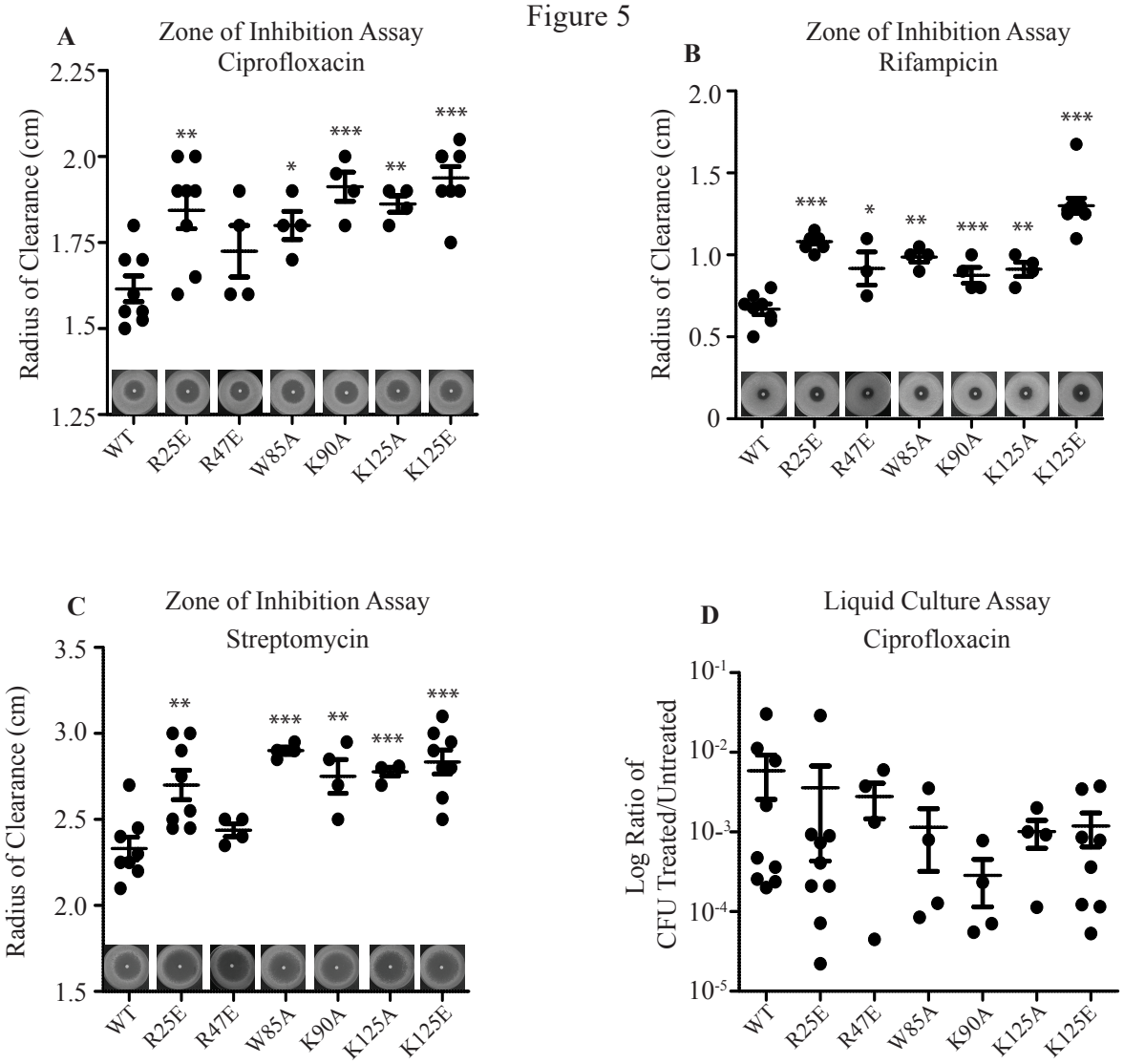

A
Zone of Inhibition Assay
Ciprofloxacin
Radius of Clearance (cm)
2.25
2.0
1.75
1.5
1.25
**
*
***
**
***
WT
R25E
R47E
W85A
K90A
K125A
K125E
B
Zone of Inhibition Assay
Rifampicin
Radius of Clearance (cm)
2.0
1.5
1.0
0.5
0
***
*
**
***
**
***
WT
R25E
R47E
W85A
K90A
K125A
K125E
C
Zone of Inhibition Assay
Streptomycin
Radius of Clearance (cm)
3.5
3.0
2.5
2.0
1.5
**
***
**
***
***
WT
R25E
R47E
W85A
K90A
K125A
K125E
D
Liquid Culture Assay
Ciprofloxacin
Log Ratio of
CFU Treated/Untreated
10-1
10-2
10-3
10-4
10-5
WT
R25E
R47E
W85A
K90A
K125A
K125E

Figure 6

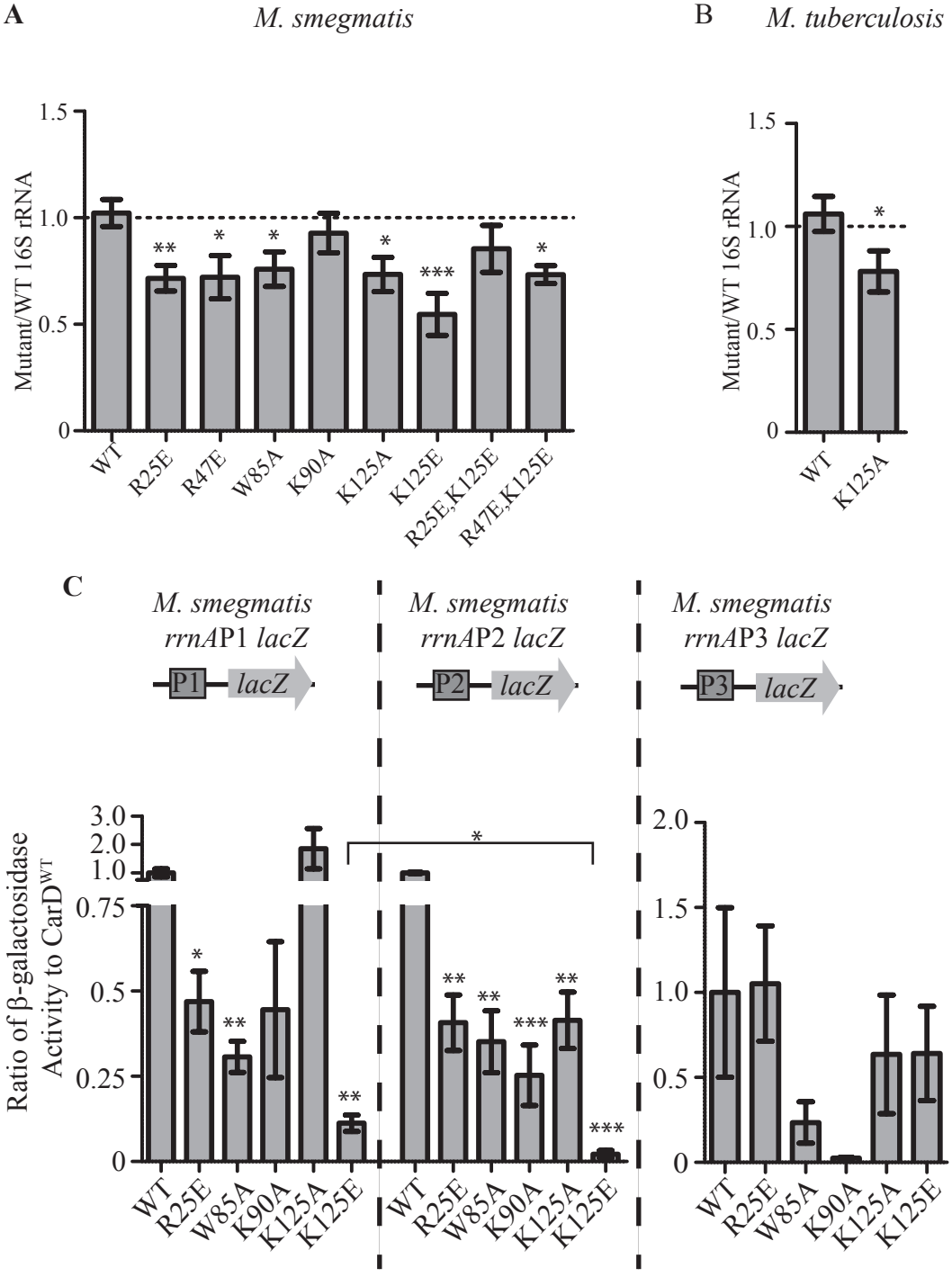

A
M. smegmatis
B
M. tuberculosis
Mutant/WT 16S rRNA
1.5
1.0
0.5
0
WT
R25E
R47E
W85A
K90A
K125A
K125E
R25E,K125E
R47E,K125E
K125A
C
M. smegmatis rrnAP1 lacZ
M. smegmatis rrnAP2 lacZ
M. smegmatis rrnAP3 lacZ
P1
P2
P3
lacZ
Ratio of β-galactosidase Activity to CarD^WT
3.0
2.0
1.0
0.75
0.5
0.25
0
1.5

Figure 7

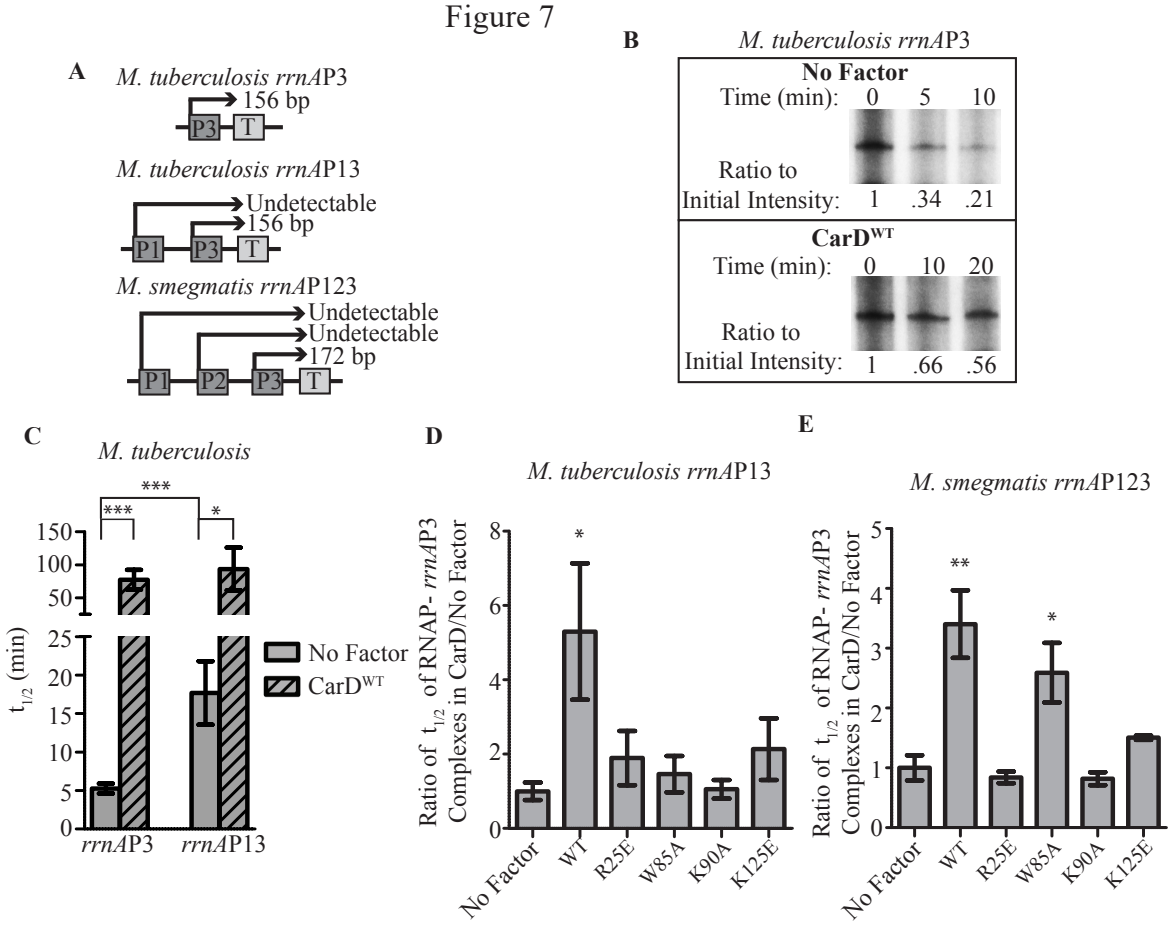
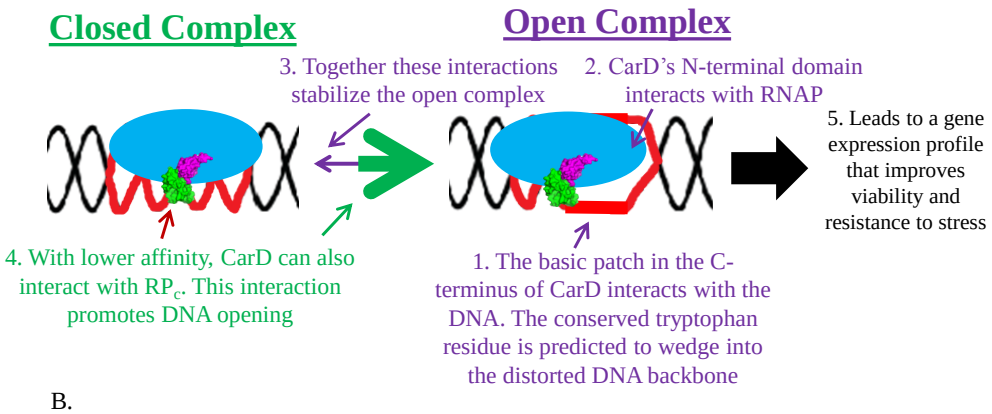
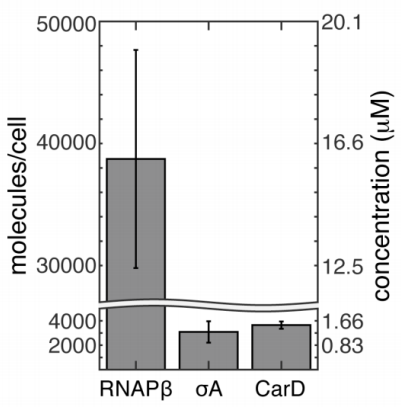

A
M. tuberculosis rrnAP3
156 bp
P3
T
M. tuberculosis rrnAP13
Undetectable
156 bp
P1
P3
T
M. smegmatis rrnAP123
Undetectable
Undetectable
172 bp
P1
P2
P3
T
B
M. tuberculosis rrnAP3
No Factor
Time (min): 0 5 10
Ratio to Initial Intensity: 1 .34 .21
CarDWT
Time (min): 0 10 20
Ratio to Initial Intensity: 1 .66 .56
C
M. tuberculosis
***
***
*
150
100
50
25
20
15
10
5
0
t1/2 (min)
No Factor
CarDWT
rrnAP3
rrnAP13
D
M. tuberculosis rrnAP13
Ratio of t1/2 of RNAP-rrnAP3 Complexes in CarD/No Factor
8
6
4
2
0
*
No Factor
WT
R25E
W85A
K90A
K125E
E
M. smegmatis rrnAP123
Ratio of t1/2 of RNAP-rrnAP3 Complexes in CarD/No Factor
5
4
3
2
1
0
**
*
No Factor
WT
R25E
W85A
K90A
K125E

Figure 8

A.


Closed Complex
Open Complex
3. Together these interactions stabilize the open complex
2. CarD's N-terminal domain interacts with RNAP
5. Leads to a gene expression profile that improves viability and resistance to stress
4. With lower affinity, CarD can also interact with RPc. This interaction promotes DNA opening
1. The basic patch in the C-terminus of CarD interacts with the DNA. The conserved tryptophan residue is predicted to wedge into the distorted DNA backbone

B.


50000
40000
30000
4000
2000
20.1
16.6
12.5
1.66
0.83
molecules/cell
concentration (μM)
RNAPβ
σA
CarD

Chapter 3: Effects of Increasing the Affinity of CarD for RNA Polymerase on *Mycobacterium tuberculosis* Growth, rRNA Transcription, and Virulence

Ashley L. Garner, Jayan Rammohan, Jeremy P. Huynh, Lucas M. Onder, James Chen, Brian Bae, Drake Jensen, Leslie A. Weiss, Ana Ruiz Manzano, Seth A. Darst, Elizabeth A. Campbell, Bryce E. Nickels, Eric A. Galburt, and Christina L. Stallings

A version of this chapter has been submitted for publication to the

Journal of Biological Chemistry

Abstract

CarD is an essential RNA polymerase (RNAP) interacting protein in *Mycobacterium tuberculosis* (*Mtb*) that stimulates formation of RNAP-promoter open complexes. CarD plays a complex role in *Mtb* growth and virulence that is not fully understood. Therefore, to gain further insight into the role of CarD in *Mtb* growth and virulence we determined the effect of increasing the affinity of CarD for RNAP. Using site-directed mutagenesis guided by crystal structures of CarD bound to RNAP, we identified amino acid substitutions that increase the affinity of CarD for RNAP. Using these substitutions, we show that increasing the affinity of CarD for RNAP increases the stability of the CarD protein in *Mtb.* In addition, we show that increasing the affinity of CarD for RNAP increases growth rate in *Mtb* without affecting ribosomal RNA (rRNA) transcription. We further show that increasing the affinity of CarD for RNAP reduces *Mtb* virulence in a mouse model of infection despite the improved growth rate *in vitro*. Our findings suggest that the CarD-RNAP interaction protects CarD from proteolytic degradation in *Mtb*, establish that growth rate and rRNA transcription can be uncoupled in *Mtb*, and demonstrate that the strength of the CarD-RNAP interaction has been finely tuned to optimize virulence.

Introduction

Mycobacterium tuberculosis (*Mtb*) remains a major global health problem, resulting in 9.6 million new cases of tuberculosis (TB) in 2014 and 1.5 million TB-related deaths that year (4). Control of the *Mtb* epidemic is hampered by the increasing prevalence of multidrug resistant strains, which resulted in 480,000 patients developing multidrug-resistant TB in 2014. In order to develop new strategies to battle this pathogen, we must gain a better understanding of the molecular processes involved in its survival and pathogenesis. Recent studies have identified

numerous aspects of mycobacterial physiology that differ from what has been identified in model organisms and highlight the importance of research directly in mycobacteria to understand their lineage specific physiology. We have identified CarD as an essential transcriptional regulator in mycobacteria that is not conserved in the model organism *Escherichia coli,* which has traditionally been used to study mechanisms of transcription (85). *carD* is conserved in all mycobacteria and numerous other bacteria (85–87, 125), but is not found in eukaryotes. Therefore, studying CarD will broaden our understanding of prokaryotic transcription while also characterizing a promising potential target for desperately needed new therapeutic strategies for TB.

Chromatin immunoprecipitation-sequencing (ChIP-seq) experiments in mycobacteria demonstrated that CarD is localized with RNA polymerase (RNAP) holoenzyme at promoters throughout the genome, indicating that CarD is a global regulator of transcription initiation (Chapter 1 Fig. 2, (87)). CarD interacts directly with the β1-lobe of the RNAP-β subunit through its N-terminal RNAP interaction domain (RID) (Chapter 2 Fig. 1 (85, 101, 125)) and with DNA just upstream of the -10 element of the promoter through a conserved basic patch in its C-terminal domain (Chapter 2 Fig. 3, (86, 87, 108)). Within the basic patch, studies have identified a highly conserved tryptophan residue that is proposed to wedge into the minor groove at the upstream edge of the transcription bubble (86, 87). Thus far, CarD's activity has primarily been studied on ribosomal RNA (rRNA) promoters. Using *in vitro* transcription assays, we have shown that CarD stabilizes RNAP-promoter open complexes at rRNA promoters during transcription initiation and that CarD requires interactions with both RNAP and DNA as well as the activity of the conserved tryptophan to do so (Chapter 2 Fig. 7, (87, 101, 108, 109)). Recently, a bulk fluorescence assay was used to measure the effect of CarD on the kinetics of

transcription initiation at the *Mtb* rRNA promoter *rrnA*P3. These studies revealed that, compared to the *E. coli* RNAP, the *Mycobacterium bovis* (*Mbo*) RNAP forms a significantly less stable open complex at *rrnA*P3 in the absence of CarD. Addition of CarD stabilizes *Mbo* RNAP-*rrnA*P3 open complexes by binding to the RNAP-promoter open complex with high affinity and preventing collapse to closed complex. With lower affinity, CarD also binds the RNAP-promoter closed complex and promotes melting of the DNA to form open complex. Importantly, the cellular concentration of CarD in wild-type (WT) *Mtb* is sufficient for both of these activities to be physiologically relevant (Chapter 3 Fig. 8, (109)). Mycobacterial strains expressing mutants of CarD with weakened affinity for the RNAP-β subunit, weakened affinity for DNA, or that are mutated at the conserved tryptophan all express lower levels of rRNA, exhibit lower rRNA promoter activation in promoter-*lacZ* fusion experiments, grow slower, and are more sensitive to multiple stresses and antibiotics as compared to strains expressing WT CarD (Chapter 3, (87, 101, 108)).

In this study, we use CarD mutants with increased affinity for RNAP to dissect the role of CarD binding to RNAP *in vitro* and *in vivo*. We detail the distinct effects of increasing the affinity of CarD to RNAP on open complex stability, transcriptional regulation, CarD protein stability, and the effects of these changes on growth and virulence. These studies expand our understanding of the global transcription regulator CarD and the impact of CarD activity on the physiology of *Mtb*, while also providing insight into how CarD activity is regulated.

Experimental Procedures

Bacterial two-hybrid assays. The bacterial two hybrid assay used in this study is based upon the demonstration that contact between a protein domain fused to the α subunit of RNA polymerase and a partner protein fused to the bacteriophage λCI protein activates transcription of

a *lacZ* reporter gene under the control of a test promoter bearing an upstream λ operator (131). Assays were performed as described using FW102 F′ O_L2-62 reporter strain cells, which contain the test promoter placO_L2-62 driving the expression of a linked *lacZ* gene on an F′ episome (131). Plasmids used in these assays included pBRα-Mt CarD (1 to 66), which encodes residues 1 to 248 of the *E. coli* RNAP α subunit fused to residues 1-66 of *Mtb* CarD (101), plasmid pBRα, which encodes the WT α subunit, and plasmid pACλCI-Mt β1, which encodes residues 1 to 236 of the bacteriophage λCI protein fused to residues 52 to 178 of the β subunit of *Mtb* RNAP fused to residues 379 to 440 of the β subunit of *Mtb* RNAP via two glycine residues (101). Plasmids carrying amino acid substitutions at residue I27 of *Mtb* CarD were introduced into pBRα-CarD (1 to 66) by PCR.

FW102 F′ O_L2-62 were cotransformed with either the indicated pBRα-Mt CarD (1 to 66) derivative and pACλCI-Mt β1 or pBRα and pACλCI-Mt β1. Individual transformants were selected and grown in LB supplemented with carbenicillin (100 μg/ml), chloramphenicol (25 μg/ml), and kanamycin (50 μg/ml). No IPTG was present in the growth media. β-galactosidase assays were performed as described in (101). Graphs in Figures 1E and 1F depict the β-galactosidase activity in cells containing the indicated pBRα-Mt CarD (1 to 66) derivative and pACλCI-Mt β1. The bar labeled "No CarD" in Figure 1E depicts the β-galactosidase activity observed in cells containing pBRα and pACλCI-Mt β1.

Bacterial strains and growth conditions. (i) *Mtb.* All *Mtb* strains were derived from the Erdman strain and were grown at 37°C in Sauton's broth media (0.5 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4$, 4 g 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) or 7H10 agar media (Difco) supplemented

with 60 µl L^{-1} oleic acid, 5 g L^{-1} bovine serum albumin (BSA), 2 g L^{-1} dextrose, 0.003 g L^{-1} catalase (OADC), and 0.5% glycerol. Gene switching was used to construct strains of mycobacteria expressing different *carD* alleles and to test for their viability (101, 110). Specifically, the *Mtb* Δ*carD attB*::*tet-carD* strain (described previously in (85)) was transformed with pMSG430*Rv3583c*^{I27F}, pMSG430*Rv3583c*^{I27W}, or pMSG430*Rv3583c*^{R47E} (expresses *Mtb* CarD^{WT}, CarD^{I27F}, CarD^{I27W}, or CarD^{R47E}, respectively, from a constitutive P*myc1-tetO* promoter, kanamycin resistant) or pDB19Rv3583c (expresses CarD^{WT} from a constitutive P*myc1-tetO* promoter, zeocin resistant). The transformants were selected on 20 µg ml^{-1} kanamycin or 1.25 µg ml^{-1} zeocin. The *carD* gene from each transformant was sequenced to confirm the presence of the correct sequence. The *Mtb* Δ*carD attB*::*tet-carD* strains transformed with pDB19*Rv3583c*^{WT}, pMSG430*Rv3583c*^{I27F}, pMSG430*Rv3583c*^{I27W}, or pMSG430*Rv3583c*^{R47E} were named mgm3080, csm230, csm231 and csm195, respectively. (ii) *M. smegmatis*. All *M. smegmatis* strains were derived from mc^2155 and were grown at 37°C in LB supplemented with 0.5% dextrose, 0.5% glycerol, and 0.05% Tween 80 (broth). The *M. smegmatis* strains expressing hemagglutinin (HA)-tagged CarD^{WT}, CarD^{R25E}, CarD^{I27F}, or CarD^{I27W} were engineered as described for the analogous *Mtb* strains using pMSG430 expression plasmids and the *M. smegmatis* Δ*carD attB::tet-carD* strain (described previously in (85)). The *M. smegmatis* Δ*carD attB*::*tet-carD* strains expressing C-terminal HA-tagged versions of *Mtb* CarD^{WT}, *Mtb* CarD^{R25E}, *Mtb* CarD^{I27F}, or *Mtb* CarD^{I27W} from a constitutive P*myc1-tetO* promoter at the *attB* site of *M. smegmatis* Δ*carD attB*::*tet-carD* were named mgm3090, mgm3091, csm228, and csm229, respectively.

Cell lysate. 20ml of exponential cultures of *Mtb* were lysed in 500 µl NP-40 buffer (10 mM sodium 182 phosphate [pH 8.0], 150 mM NaCl, and 0.25% NP-40) by bead beating three

times.

Immunoprecipitation. Lysate from *M. smegmatis* strains expressing HA-tagged alleles of CarD was bound to monoclonal anti-HA agarose (Sigma) and complexes were eluted as previously described (101, 108).

Western blot analysis. Protein samples were mixed with an SDS-PAGE loading buffer and ran on a 4-12% Bis-Tris protein gel (Invitrogen). HA-tagged CarD in the immunoprecipitation experiment was detected using a mouse monoclonal antibody (clone 10F05; Memorial Sloan-Kettering Cancer Center Monoclonal Antibody Core Facility). To determine protein concentrations in cell lysate, CarD was detected with a rabbit polyclonal antibody (85). RNAP-β was detected with a mouse monoclonal antibody in both experiments (clone 8RB13; Neoclone, Madison, WI). Horseradish peroxidase conjugated secondary antibodies against mouse or rabbit (Perkin Elmer) were used for detection.

Native gel electrophoretic mobility shift assays (EMSAs) by radiolabeling. A DNA fragment containing the *rrnA* promoter and leader sequences, called *rrnA*PL and corresponding to *M. smegmatis* mc^2155 nucleotides 5,029,577–5,029,909, was prepared and radiolabeled as previously described (87). 1 pmol of labelled DNA was incubated with 200 pmol of *M. smegmatis* $CarD^{WT}$, *M. smegmatis* $CarD^{K125E}$, *Mtb* $CarD^{I27F}$, or *Mtb* $CarD^{I27W}$ proteins for 20 min at room temperature and analyzed via electrophoresis as previously described (87).

RNA was prepared from 20-30 ml of log-phase *Mtb* Erdman (WT *Mtb*), mgm3080 ($CarD^{WT}$), csm195 ($CarD^{R47E}$), csm230 ($CarD^{I27F}$), and csm231 ($CarD^{I27W}$). 16S rRNA and *carD* levels were measured and normalized to *sigA* transcript levels as previously described (85).

Mouse infections. Infection of mice with exponentially replicating *Mtb* Erdman strain (WT *Mtb),* csm3080 ($CarD^{WT}$), csm230 ($CarD^{I27F}$) and csm231 ($CarD^{I27W}$) strains and

determination of bacterial loads was performed as previously described (101, 108). All procedures involving animals were conducted according to the National Institutes of Health (NIH) guidelines for the housing and care of laboratory animals, and they were performed in accordance with institutional regulations after protocol review and approval by the Institutional Animal Care and Use Committee of the Washington University in St. Louis School of Medicine (protocol 20130156, Analysis of Mycobacterial Pathogenesis). Washington University is registered as a research facility with the United States Department of Agriculture and is fully accredited by the American Association of Accreditation of Laboratory Animal Care. The Animal Welfare Assurance documentation is on file with the Office for Protection from Research Risks of the NIH. All animals used in these experiments were subjected to no or minimal discomfort. All mice were euthanized by CO_2 asphyxiation, which is approved by the American Veterinary Medical Association Panel on Euthanasia.

Protein preparation for biochemical assays. Recombinant *Mycobacterium bovis* (*Mbo*) core RNAP was purified from *E. coli* using a system kindly supplied by Dr. Robert Landick (113, 114) as previously described (108). Recombinant *Mbo* σ^A, which is identical to *Mtb* σ^A, was purified from *E. coli* and added to the core RNAP to reconstitute the RNAP holoenzyme.

Preparation of fluorescent promoter DNA fragments. The DNA template contains nucleotides 1,470,151 to 1,470,300 of the *Mtb* Erdman genomic DNA, which includes the *rrnA*P3 promoter. Cy3-labeled promoter DNA was prepared as previously described (109).

Stopped-flow fluorescence assay. Stopped-flow fluorescence experiments were performed as previously described (109). All experiments were performed using a final *Mbo* RNAP-σ^A holoenzyme concentration of 225 nM and a final promoter DNA concentration of 10 nM. Experiments were performed at 25° C in the following final solution conditions: 14 mM

Tris pH 8.0, 120 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, and 10% glycerol by volume. Equal volume mixing of protein with promoter DNA was performed in a stopped-flow apparatus (Applied Photophysics SX-20, total shot volume 150 µl, dead time < 2 ms), so the initial protein and DNA solutions were each diluted by half in order to reach their final reaction concentrations. Due to slight hardware variations in different stopped-flow instruments, excitation light was provided by either a 510 nm fixed-wavelength LED light source or a 515 nm light source from an arc lamp passed through a monochromator. The difference in excitation light had no effect on the data. In all cases emission was collected at 570+ nm using a long-pass filter. Fluorescence was monitored for 20 min.

At least 2 traces were collected and averaged per condition. Traces were plotted as fold-change over DNA according to (F-F0)/F0, where F0 is the buffer-subtracted signal for DNA alone and F is the buffer-subtracted signal for DNA mixed with protein. The final point of each trace was used as a measure of equilibrium fluorescence. The fold-change traces were fit to a triple exponential from 0.1-1200 s using the ProData Viewer software from Applied Photophysics. The slowest observed rate dominated the fractional amplitude of the fits and was used as a measurement of k_{obs}[3]. Conditions that were repeated multiple times on different days were used to estimate standard error of the mean (SEM). An average SEM was used to estimate uncertainty for specific conditions that were only repeated multiple times on the same day.

***In vitro* transcription assays.** CarD proteins used in *in vitro* transcription assays were diluted into 1× dialysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, and 1 mM BME). For the aborted 3nt transcript assay (123), a linear fragment of dsDNA *Mtb* Erdman strain genomic DNA containing nucleotides 1,470,151 to 1,470,300 which includes the *Mtb rrnA*P3 promoter was prepared by annealing and extending primers overlapping 85bp primers (Integrated DNA

Technologies Coralville, USA). Reaction conditions were as follows: 200 nM *Mbo* core RNAP, 2 μM *Mbo* σ^A, 2 μM CarD or equivalent volume of buffer, 10 nM linear DNA template, 210 μM GpU dinucleotide, 21 μM UTP, 0.1 μl [α - 32 P]-UTP, 13.25 mM Tris pH 8.0, 59 mM NaCl, 10.12 mM $MgCl_2$, 5% (vol/vol) glycerol, 1mM DTT, and 0.1 mg/ml BSA (NEB) in a total volume of 20 μl. *Mbo* core and σ^A were incubated for 10 min at 37°C followed by the addition of CarD and 10 more minutes of incubation at 37°C. The DNA template was added and the reactions were diluted to 17.5 μl and incubated at 37°C for 10 minutes. Reactions were initiated with addition of a 2.5 μl mixture containing GpU, UTP, and the radiolabeled UTP and incubated at 37°C. After 20 min, the reactions were stopped with 2 × formamide buffer [98% (vol/vol) formamide, 5 mM EDTA] and run on a 22% urea PAGE gel. Transcripts were quantified using phosphorimagery and analyzed using Image Gauge software.

Results

Identification of amino acid substitutions in the N-terminus of CarD that increase the affinity of CarD for RNAP-β. To identify substitutions that strengthen the *Mtb* CarD/RNAP-β β1-lobe interaction, we employed site directed mutagenesis guided by a comparison between our X-ray structure of *Thermus thermophilus* (*Tth*) CarD in complex with the β1-lobe of *Tth* RNAP, solved to 2.4 Å resolution (4XAX (86); Fig. 1A), and the structure of *Mtb* CarD in complex with the RNAP-β β1 and β2 lobes of *Mtb* RNAP (4KBM; (132)) solved to 2.11 Å. The β1-lobes from each structure were aligned over 919 atoms in PyMOL (version 1.8 Schrödinger, LLC) to give an RMS of 1.095 Å and interactions between the β1-lobe and the CarD-RID domains were compared. Many of the interactions were conserved between the *Tth* and *Mtb* structures (Fig. 1, B and D), but a significant difference was noted in a region that included a loop at the tip of the RID (residues 26-29 in *Mtb*) that was not modeled in the *Mtb*

structure, presumably due to a lack of density. Cartoon rendering by B-factors (PyMOL) of both structures revealed that this loop was highly disordered in *Mtb*, but well-ordered in *Tth*, presumably due to its interaction with the β1-lobe (Fig. 1C). In particular, the *Tth* structure revealed a non-polar interaction between loop residue V28 from CarD (I27 in *Mtb*) and I101 and I108 from the β1-lobe (I140 and I147 in *Mtb*) that was not visible in the *Mtb* structure due to disorder (Fig. 1B). Based on the *Tth* structure of the β1-lobe in complex with CarD, we predicted that the isoleucine side chain at position 27 of *Mtb* CarD likely contributes to an interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe that may not have been observed in the crystal structure due to crystal packing constraints. Furthermore, structural modeling predicts that removal of the isoleucine side chain at position 27 of the *Mtb* CarD-RID would weaken the interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe whereas substitutions of position 27 with residues containing side chains that are more hydrophobic than isoleucine might strengthen the interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe.

To test these predictions, we employed a bacterial two-hybrid assay that detects the interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe (101). In this assay, the interaction between WT *Mtb* CarD-RID and the *Mtb* β1-lobe results in a 3-fold increase in β-galactosidase (β-gal) activity from the two-hybrid test promoter (Fig. 1E). Substitution of I27 of the CarD-RID with alanine or glycine reduced the β-gal activity to near background levels while substitution of I27 of the CarD-RID with the bulkier hydrophobic residues phenylalanine and tryptophan resulted in a 7-fold and 9-fold increase, respectively, in β-gal activity relative to that observed with the WT *Mtb* CarD-RID (Fig. 1, E and F). The results of the two hybrid assays indicate that I27A and I27G substitutions weaken the interaction between the *Mtb* CarD-RID and the *Mtb* β1-

lobe while I27F and I27W substitutions strengthen the interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe, consistent with the predictions from the structural modeling.

To determine if the CarD I27F and I27W mutations increase the affinity of the interaction between CarD and RNAP-β subunit *in vivo*, we constructed *Mycobacterium smegmatis* strains expressing HA-tagged $CarD^{WT}$, $CarD^{R25E}$, $CarD^{I27F}$, or $CarD^{I27W}$. Co-immunoprecipitation experiments in *M. smegmatis* demonstrated that $CarD^{I27W}$ and $CarD^{I27F}$ mutants co-precipitated more RNAP-β than $CarD^{WT}$ (Fig. 1G). As previously reported, the $CarD^{R25E}$ mutant, which has lower affinity for RNAP-β than $CarD^{WT}$, co-precipitated less RNAP-β than $CarD^{WT}$ (101). The $CarD^{I27W}$ mutant associated with more RNAP-β than $CarD^{I27F}$, indicating a higher affinity interaction. To determine if mutations at I27 specifically affect only the interaction with RNAP and not DNA, we performed an electrophoretic mobility shift (EMSA) assay previously used to demonstrate CarD's DNA binding activity (87, 108). This assay uses the *rrnA*PL DNA fragment, which contains the promoters and leader sequences of the *M. smegmatis* rRNA *rrnA* operon, and was chosen due to CarD's described role in regulating rRNA transcription. In contrast to a mutation in the DNA binding domain of CarD (K125E) that abolished the ability of CarD to bind and shift DNA (108), neither $CarD^{I27F}$ nor $CarD^{I27W}$ affected CarD's ability to bind and shift DNA, indicating that their effects are limited to CarD's interaction with RNAP (Fig. 1H).

$CarD^{I27F}$ and $CarD^{I27W}$ mutants stabilize open complexes at lower concentrations than $CarD^{WT}$. CarD regulates transcription initiation by binding to and stabilizing RNAP-promoter open complexes (Chapter 3, (86, 87, 108, 109, 123)). We have previously developed a fluorescence assay for CarD activity that reports on open complex formation in real-time. In this assay, a Cy3 label, which exhibits a 2-fold fluorescence enhancement in open complex, is incorporated at the +2 position on the non-template strand of the *Mtb rrnA*P3 promoter within a

linear fragment of *Mtb* genomic DNA containing nucleotides 1,470,151 to 1,470,300 (109). *Mbo* RNAP-σ^A holoenzyme with or without *Mtb* CarD is then mixed with the labeled DNA fragment via stopped-flow spectrophotometry and fluorescence is monitored for 20 minutes. The Cy3 label increases fluorescence in the open complex conformation and thus the rate of open complex formation can be monitored as a change in fluorescence. The amplitude of the fluorescence intensity curve correlates to the equilibrium amount of open complex in a given condition (109). To monitor the effect of increasing the affinity of CarD for RNAP on the formation and stability of open complexes, we performed this assay with concentrations of $CarD^{I27F}$ and $CarD^{I27W}$ ranging for 0 to 1800 nM and fixed concentrations of RNAP (225 nM) and promoter DNA (10 nM) (Fig. 2A). We found that the CarD concentration necessary to reach half of the maximum level of open complex at saturation (half-maximal concentration) was lower for $CarD^{I27F}$ (17 ± 2 nM) and $CarD^{I27W}$ (23 ± 3 nM) than for $CarD^{WT}$ (59 ± 10 nM), consistent with a higher affinity of the CarD I27 mutants for RNAP in initiation complexes compared to $CarD^{WT}$ (Fig. 2A).

$CarD^{I27F}$ and $CarD^{I27W}$ mutants accelerate promoter opening at lower concentrations than $CarD^{WT}$. We have previously reported modeling based on the trends of the slowest observed rate in the real-time fluorescent traces (k_{obs}^3) that suggests that CarD stabilizes open complex through a two-tiered concentration-dependent mechanism where CarD associates with both open and closed complexes with different affinities (109). More specifically, the model predicts that at low concentrations (i.e. < 100 nM), $CarD^{WT}$ binds to open complex and prevents bubble collapse, resulting in more open complex and a slower observed rate. The model further predicts that at higher concentrations, $CarD^{WT}$ binds to the closed complex and accelerates the rate of opening, resulting in still more open complex and an acceleration in the observed rate. Analysis of k_{obs}^3 as a function of CarD concentration was performed for the CarD I27 mutants as

previously described (109) (Fig. 2B). We found that the I27 mutants begin to accelerate k_{obs}^3 at lower concentrations (< 50 nM) than WT CarD (100 nM). In fact, for $CarD^{I27W}$ we did not observe a deceleration in k_{obs}^3 even at the lowest concentration tested (16 nM). The two-tiered kinetic model predicts that acceleration in k_{obs}^3 arises from CarD binding to closed complex, which increases the rate of promoter opening. Thus, our working model predicts that, even at 16 nM, $CarD^{I27W}$ is associating with closed complex and accelerating opening. The acceleration of the observed rate at lower concentrations coupled with the lower half-maximal concentration for open complex stabilization is consistent with a model where the CarD I27 mutants have higher affinities to both closed and open RNAP-promoter complexes as compared to $CarD^{WT}$.

Increasing the affinity of CarD to RNAP increases formation of initial RNA products. To investigate whether increasing the affinity of CarD for RNAP affects RNA synthesis, we monitored initial product formation in a multi-round *in vitro* transcription assay (123). *Mbo* RNAP-σ^A holoenzyme was incubated with the promoter template used in the fluorescence assays, but without the Cy3 label, in the presence or absence of WT or mutant *Mtb* CarD. GpU dinucleotide was used as the initiating nucleotide and radiolabeled uridine triphosphate (UTP) was added as the extending nucleotide. Reactions were incubated for 20 minutes, after which the amount of initial trinucleotide RNA products (GpUpU) were quantified. Addition of $CarD^{WT}$ protein increases the amount of initial RNA product formation ~6 fold as compared to the amount of product produced in the absence of CarD (Fig. 2C), whereas $CarD^{R25E}$, which has decreased affinity for RNAP, only increases the amount of product formed ~2 fold. In contrast, $CarD^{I27F}$ and $CarD^{I27W}$ increase the amount of product formed ~9 fold. These data demonstrate that increasing the affinity of CarD to the RNAP results in an increase in initial RNA product formation. We infer that the effect of strengthening the interaction between

CarD and RNAP on initial RNA product formation is a consequence of an increase in the formation of RNAP-promoter complexes (Fig 2).

Increasing the affinity of CarD for RNAP stabilizes the CarD protein in *Mtb*. In previous work, we constructed a strain of *Mtb* in which a copy of *carD* is expressed from a constitutive promoter at the *attB* site, a non-endogenous chromosomal locus, and the endogenous copy of *carD* is deleted (85). We have previously shown that a strain of *Mtb* that expresses $carD^{R47E}$, which encodes a protein with weakened affinity to the RNAP, from the *attB* site results in slower growth than a strain expressing $carD^{WT}$ from the *attB* site (Chapter 1, Fig. 1, (101)). To determine the effects of increasing the affinity of CarD for RNAP in the bacteria, we constructed *Mtb* strains expressing $carD^{I27F}$ or $carD^{I27W}$ from the *attB* site. We first determined the expression levels of the *carD* alleles in each strain by measuring carD transcript and protein levels. We found the $CarD^{WT}$, $CarD^{I27F}$, $CarD^{I27W}$, and $CarD^{R47E}$ all produce equal levels of *carD* transcript, as expected given that the strains express the *carD* gene from the same promoter (Fig. 3A). However, when we examined protein content in these strains we found that the $CarD^{I27F}$ and $CarD^{I27W}$ strains reproducibly had more CarD protein than the $CarD^{WT}$ strain (Fig. 3B). Additionally, this effect is proportional to the magnitude of the change in affinity, as the $CarD^{I27W}$ strain reproducibly has more CarD protein than the $CarD^{I27F}$ strain and the $CarD^{WT}$ strain had more CarD protein than $CarD^{R47E}$. Given the equal amount of transcript being produced, the different levels of protein indicate post-translational differences in the stability of CarD protein in these strains. CarD has previously been identified as a Clp protease substrate in *Mtb* (133). Taken together, our data can be explained in the context of a model where CarD is protected from proteolysis when associated with RNAP.

Increasing CarD's affinity for RNAP increases growth rate in *Mtb*. We have previously reported that decreasing CarD's affinity for RNAP decreases growth rate in *Mtb* strain (Chapter 1 Fig 1, (101)). To determine the effect of increasing CarD's affinity for RNAP on growth rate, we monitored growth in liquid Sauton's media and found that the $CarD^{I27F}$ and $CarD^{I27W}$ strains had a faster doubling time than the $CarD^{WT}$ strain (26.31, 25.75, and 30.74 hours doubling time, respectively) (Fig. 3, C and D). These results, as well as the growth defect previously reported in a strain expressing a CarD mutant with lower affinity for RNAP (101), demonstrate that the affinity of CarD to the RNAP can impact growth rate in *Mtb*.

Increasing the affinity of CarD to RNAP does not result in differences in rRNA transcript levels in *Mtb*, despite the effects on growth rate. We have previously shown that decreasing the affinity of CarD for RNAP results in decreased rRNA levels and slower growth (Chapter 1 Fig. 1(101)). Because rRNA levels are often correlated to growth rate (119, 134, 135), the decreased rRNA levels in the CarD mutants with lower affinity for RNAP could contribute to the growth defect associated with these mutations. To determine the effect of increasing the affinity of the interaction between CarD and RNAP on regulation of rRNA transcription in the bacterium, we isolated RNA from exponential cultures of $CarD^{WT}$, $CarD^{I27F}$, and $CarD^{I27W}$ strains in Sauton's media and performed quantitative real-time PCR (qRT-PCR) analysis for the levels of 16S rRNA (Fig. 3E). The CarD I27 mutants that have higher affinity for RNAP express the same amount of rRNA as the $CarD^{WT}$ strain, indicating that the faster growth in these strains is not a result of increasing rRNA expression. Therefore, rRNA levels and growth rate are uncoupled in strains expressing CarD proteins with varying affinities to RNAP. Thus, the effects of altering CarD's affinity to RNAP on growth rate can occur independently of effects on rRNA

transcription. To the best of our knowledge, this is also the first report of rRNA transcription and growth rate being uncoupled in *Mtb*.

Increasing the affinity of CarD for RNAP attenuates survival of *Mtb* during infection of mice. We have previously shown that weakening the interaction of CarD with RNAP results in decreased bacterial survival by 35 days post infection (dpi) in mice (Chapter 1, Fig. 1, (101)). To determine the effect of increasing the affinity of CarD for RNAP on virulence, we infected C57BL/6J mice with the *Mtb* $CarD^{WT}$, $CarD^{I27F}$, or $CarD^{I27W}$ strains and measured bacterial burden in the lungs at 21 and 35 dpi (Fig. 4). Both the $CarD^{I27F}$ and $CarD^{I27W}$ mutants were attenuated for survival in the mice and displayed significantly lower bacterial burden in the lungs (Fig. 4). Specifically, the $CarD^{I27W}$ strain displayed significantly lower bacterial burden than the $CarD^{WT}$ strain at 21 and 35 dpi while the $CarD^{I27F}$ strain had lower bacterial burdens than $CarD^{WT}$ at 35 dpi. The severity of the virulence defect in the CarD mutants mirrors the effect of these mutations on the affinity of CarD for RNAP, with $CarD^{I27W}$ exhibiting both a higher affinity and a more pronounced virulence defect than $CarD^{I27F}$. These data show that although increasing the affinity of CarD to RNAP increases the rate of formation of transcription competent RNAP-promoter complexes and increases growth rates in culture, it results in a loss of virulence. The degree of attenuation at 35 dpi for the CarD I27 mutants is similar to that previously reported for *Mtb* strains expressing a mutant CarD protein with lower affinity to the RNAP (101). Thus, either decreasing or increasing the affinity of CarD for RNAP has a detrimental effect on virulence of *Mtb*. We, therefore, propose that this indicates that the affinity of CarD for RNAP is finely tuned to the physiology of this obligate pathogen to optimize virulence.

Discussion

In this manuscript, we have investigated both the biochemical and physiological effects of increasing the affinity of CarD for RNAP. Our findings reveal new aspects of CarD regulation and the effect of increasing the affinity of CarD for RNAP-β on growth rate and virulence.

Using a real-time fluorescence assay we found that increasing the affinity of CarD to RNAP leads to a decrease in the concentration of CarD required for half-maximal open complex formation and stability (Fig 2A). Our data further show that the higher affinity CarD mutants result in increased stability of an RNAP-promoter complex and an increase in the formation of initial RNA products (Fig. 2C). However, despite these positive effects of the CarD I27 mutants on the formation of transcription competent RNAP-promoter complexes at *rrnA*P3 *in vitro*, *Mtb* strains expressing $CarD^{I27F}$ or $CarD^{I27W}$ did not contain higher levels of rRNA than the $CarD^{WT}$ strain (Fig. 3E). Therefore, there must be other factors at play *in vivo* that impact whether rRNA transcription responds to changes in the affinity of CarD to the RNAP.

We found that increasing to affinity of CarD for RNAP results in an increase in growth rate without changing the levels of rRNA (Fig. 3C and D). Furthermore, we found that increasing the affinity of CarD to RNAP in *Mtb* results in attenuated virulence in mice (Fig. 4). Together these findings demonstrate that the differences in growth rate and virulence of the CarD I27 mutant strains are not a result of altered rRNA content but are instead another consequence of varied levels or activity of CarD proteins, thus uncoupling the role of CarD in regulating rRNA content from its effect on growth rate. Therefore, the essential function of CarD is unlikely to be limited to regulation of rRNA expression. In fact, CarD is localized to promoter-bound RNAP holoenzymes throughout the genome (87). We have shown that in the absence of CarD, the mycobacterial RNAP is only weakly able to form open complexes at *rrnA*P3 (86, 109, 123).

These studies suggest that CarD may in fact be a general component of the transcription initiation machinery in mycobacteria and that its essential role is enabling the mycobacterial RNAP to efficiently form open complexes at all promoters in the genome. In particular, changes in the affinity of CarD to the RNAP could result in a global dysregulation of gene expression. Alternatively, phenotypes of CarD mutants may result from deregulation of a specific subset of genes differentially regulated by CarD. It should also be noted that, despite the presence at all promoters, at this point CarD activity has only been evaluated at mycobacterial rRNA promoters and consensus promoters from *E. coli*. Therefore, a priority of future work will be to clarify the role of CarD in global transcription regulation.

Using multiple CarD mutants, we have also revealed a role for the CarD-RNAP interaction in regulating CarD protein levels. Increasing the affinity of CarD for RNAP increases the cellular concentration of CarD without affecting *carD* transcript levels (Fig. 3, A and B). The effect on CarD protein concentration is proportional to the affinity for RNAP and, of the strains expressing *carD* from the *attB* site, CarD^{I27W} has the highest affinity for RNAP and the highest concentration of CarD protein while CarD^{R47E} has the lowest affinity for CarD and the lowest CarD protein levels (Fig. 3, A and B). These data suggest that CarD is protected from proteolytic degradation while associated with RNAP. Since CarD is a known target of the Clp protease in mycobacteria (133), one possibility is that Clp-mediated degradation of CarD is more efficient when CarD is not associated with RNAP. Thus, our findings provide a mechanistic explanation for our previous observation that mycobacteria contain a similar concentration of CarD and the housekeeping sigma factor (109) and suggest that there may be a detrimental effect of excess, free cellular CarD.

Acknowledgements

The authors would like to thank Dr. Tim Lohman for the use of his stopped-flow spectrophotometer, Dr. Tom Ellenberger for the use of the HPLC and FPLC in his laboratory, and Yocheved Gensler for assistance with cloning. We are grateful to Katherine Mann and Jerome Prusa for their thoughtful review of this manuscript. C.L.S. and E.A.G. are supported by Grant GM107544 from the National Institutes of Health. B.E.N. is supported by grant GM118059 from the National Institutes of Health. E.A.C. is supported by grant GM114450 from the National Institutes of Health. A.L.G. is supported by the NIGMS Cell and Molecular Biology Training Grant GM007067 and the Stephen I. Morse Graduate Fellowship. J. R. is supported by a Sigma Aldrich Predoctoral Fellowship. J.P.H. is supported by a National Science Foundation Graduate Research Fellowship DGE-1143954.

Figure Legends

Figure 1. Single point mutations in CarD specifically increase CarD's affinity to RNAP-β.

A. Crystal structure of *Tth* CarD in complex with the β1-lobe of *Tth* RNAP reveals molecular interactions between the two proteins at high resolution (2.4 Å) detail.

B. Structural alignment of the β1-lobes of *Tth* CarD:β1-lobe and *Mtb* CarD:β1-lobe structures to compare the interactions between CarD-RID and RNAP-β1 in each organism.

C. Atom displacement was rendered visually using PyMOL B-factor rendering. The spectrum scale above illustrates a range of low B-factors (highly ordered, blue indicates under 30%) to high B-factors (disordered, red indicates 90% or higher).

D. Alignment of *Tth* (top) and *Mtb* (bottom) CarD-RID comparing interactions with their cognate RNAP β1-lobe. Black shading indicates residues that are identical and yellow

shading indicates residues that are homologous. The following groups are considered homologous: (R, K), (E, D), (V, I, L, M, A), (S, T), and (Q, N). Bullets are defined in the figure by interactions unique to each organism or shared by both.

E-F. Results of β-galactosidase assays (101, 131) examining the effects of amino acid substitutions at position 27 of the *Mtb* CarD-RID on the interaction between the *Mtb* CarD-RID and the *Mtb* β1-lobe. The bar graphs show β-galactosidase activity in Miller units observed in assays done using cells containing the indicated *Mtb* CarD-RID derivative and the *Mtb* β1 lobe as well as assays done using cells containing only the *Mtb* β1 lobe. Each graph shows the mean ± SEM of data from five replicates. Statistical significance was analyzed by ANOVA and Tukey's multiple comparison test. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; or ****, $p \leq 0.0001$.

G. Immunoprecipitation experiments with a monoclonal antibody specific for HA in the *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$-HA (lane 1), $CarD^{R25E}$-HA (lane 2), $CarD^{I27F}$-HA (lane 3), or $CarD^{I27W}$-HA (lane 4). Immunoprecipitated eluates were analyzed by western blotting with monoclonal antibodies specific for either RNAP-β or CarD. *(H)* Autoradiograph of EMSA with *Mtb* $CarD^{I27F}$, *Mtb* $CarD^{I27W}$, no protein, *M. smegmatis* $CarD^{K125E}$, or *M. smegmatis* $CarD^{WT}$ incubated with 20,000 cpm (approximately 0.6 ng) of [γ32P]ATP-radiolabeled *M. smegmatis rrnA*PL DNA. The reactions were separated on a non-denaturing polyacrylamide gel, which was then dried and exposed to film.

Figure 2. Effects of increasing the affinity of CarD for RNAP on open complex stability.

A. Equilibrium fluorescence fold-change normalized to the fold-change at high CarD concentrations (saturation) for each mutant of CarD with 225 nM *Mbo* RNAP-σ^A and

10 nM Cy3-labeled *Mtb rrnA*P3 promoter. Both *Mtb* CarD^{I27F} (red triangles) and CarD^{I27W} (blue squares) mutants achieve a half-maximal effect (dashed line) at lower concentrations (17 ± 2 nM for CarD^{I27F}, and 23 ± 3 nM for CarD^{I27W}) than *Mtb* CarD^{WT} (black circles, 59 ± 10 nM).

B. k_{obs}[3], calculated using ProData Viewer software from Applied Photophysics, of open complex formation as a function of CarD concentration for *Mtb* CarD^{WT} (black circles), CarD^{I27F} (red triangles), and CarD^{I27W} (blue squares). The legend for *B* is shared with *A*.

C. *In vitro* transcription assay showing a representative gel and a graph of the ratio of the amount of 3nt initial transcript formed by 200 nM *Mbo* RNAP-σ^A from 10 nM of a linear DNA fragment in the presence of 2 μM CarD versus in the absence of CarD for reactions containing no CarD, *Mtb* CarD^{WT}, CarD^{R25E}, CarD^{I27F}, or CarD^{I27W}. The graph shows the mean ± SEM of data from at least four replicates. Statistical significance was analyzed by ANOVA and Tukey's multiple comparison test. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; or ****, $p \leq 0.0001$.

Figure 3. Increasing CarD's affinity for RNAP results in more stable CarD protein and faster growth without a change in rRNA expression.

A. The ratio of *carD* transcript levels in exponential cultures of the *ΔcarD attB::tet-carD* strain expressing CarD^{I27F}, CarD^{I27W}, or CarD^{R47E} to levels in the CarD^{WT} strain when grown in Sauton's media. Transcript levels were determined using qRT-PCR and normalized to *sigA*.

B. Western blot analysis of two biological replicates of lysates from the same cultures as used in (A). Membranes were blotted with a monoclonal antibody against RNAP-β (top) or a polyclonal antibody against CarD (bottom).

C. Representative growth curves of the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ (circles), $CarD^{I27F}$ (squares), or $CarD^{I27W}$ (triangles) in Sauton's media.

D. Doubling times of the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$, $CarD^{I27F}$, or $CarD^{I27W}$ in Sauton's media.

E. Ratio of 16S rRNA levels in exponential cultures of the *ΔcarD attB::tet-carD* strain expressing $CarD^{I27F}$ or $CarD^{I27W}$ to levels in the $CarD^{WT}$ strain when grown in Sauton's media. Transcript levels were determined using qRT-PCR and normalized to *sigA*.

Each graph shows the mean ± SEM of data from at least three replicates. Statistical significance was analyzed by ANOVA and Tukey's multiple comparison test. n.s. = not significant, *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; or ****, $p \leq 0.0001$.

Figure 4. Increasing CarD's affinity for RNAP attenuates *Mtb* in a mouse model of infection.

Dot plot of colony forming units (CFU) in the lungs of C57BL/6J mice infected with the *ΔcarD attB::tet-carD* strain expressing either $CarD^{WT}$ (circles), $CarD^{I27F}$ (squares), or $CarD^{I27W}$ (triangles). Each time point is the mean ± SEM of data from at least 6 mice per strain. Statistical significance was analyzed by ANOVA and Tukey's multiple comparison test. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, or $p \leq 0.001$. All comparisons were tested and only significantly different comparisons are shown.

Figure 1

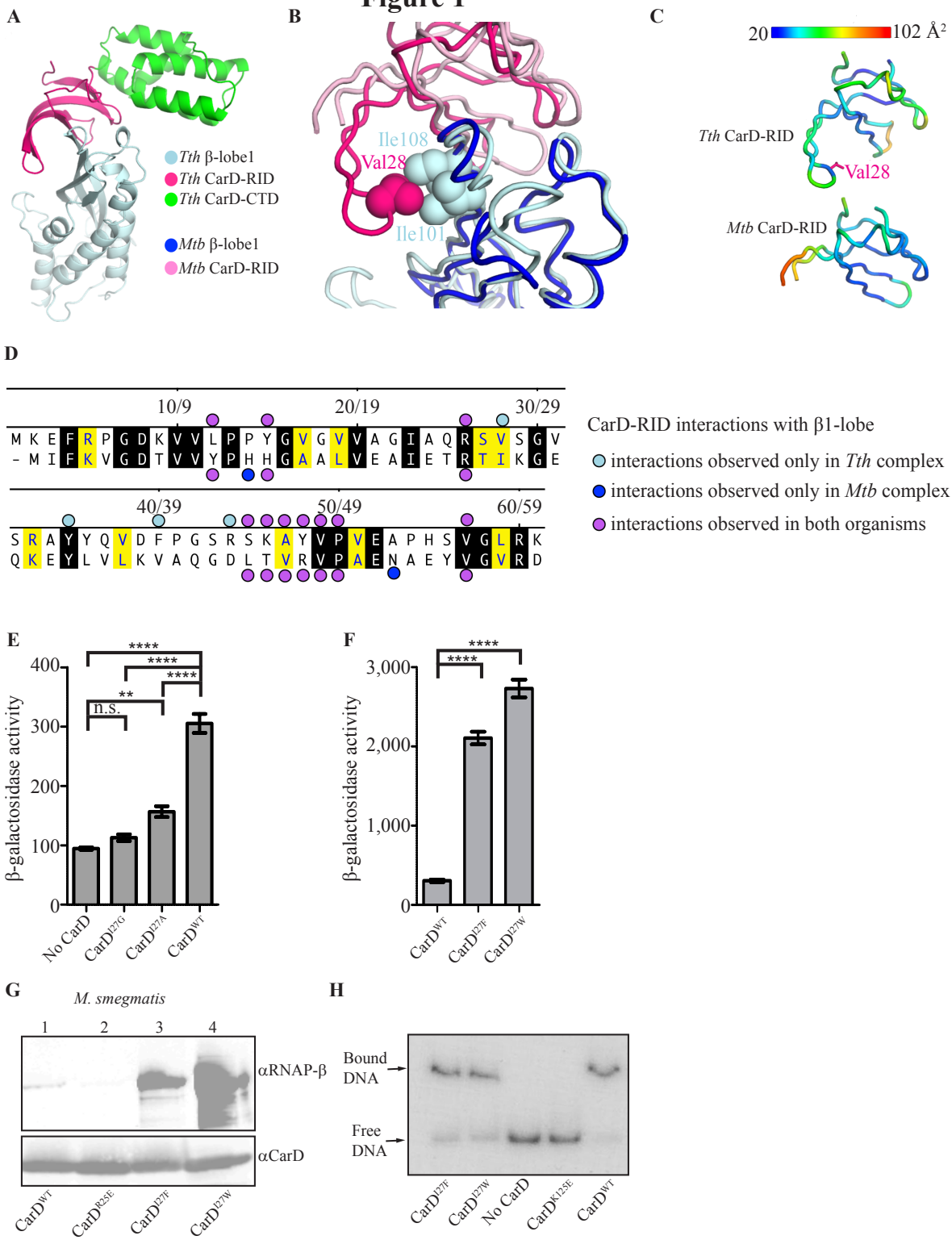
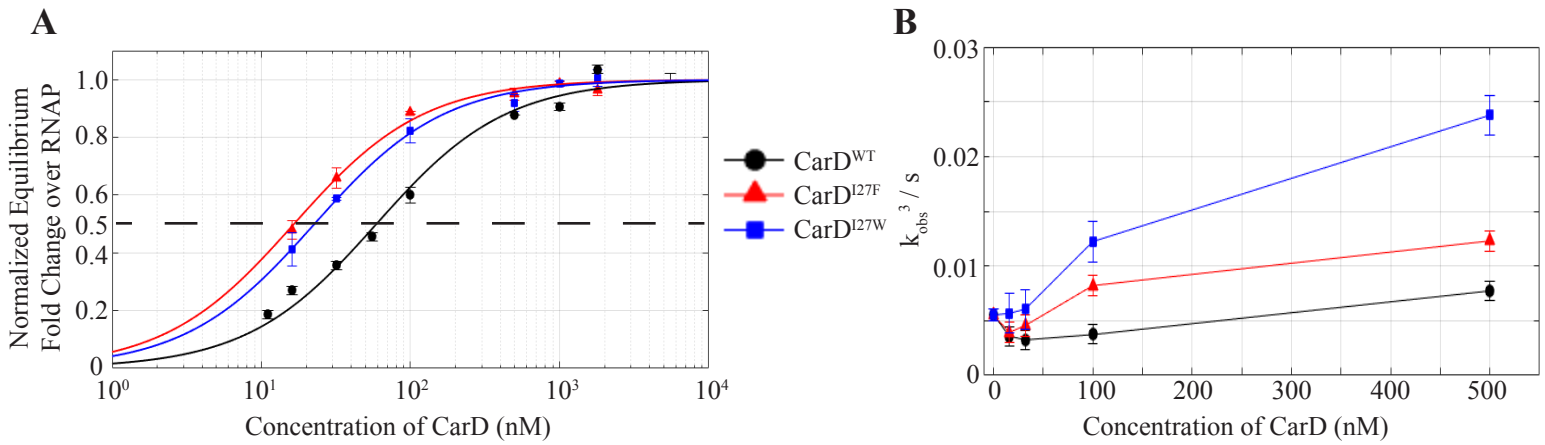

A
Tth β-lobe1
Tth CarD-RID
Tth CarD-CTD
Mtb β-lobe1
Mtb CarD-RID
B
Ile108
Val28
Ile101
C
20
102 Å²
Tth CarD-RID
Val28
Mtb CarD-RID
D
10/9
20/19
30/29
40/39
50/49
60/59
CarD-RID interactions with β1-lobe
interactions observed only in Tth complex
interactions observed only in Mtb complex
interactions observed in both organisms
E
β-galactosidase activity
n.s.
No CarD
CarD^I27G
CarD^I27A
CarD^WT
F
β-galactosidase activity
CarD^WT
CarD^I27F
CarD^I27W
G
M. smegmatis
αRNAP-β
αCarD
CarD^WT
CarD^R25E
CarD^I27F
CarD^I27W
H
Bound DNA
Free DNA
CarD^I27F
CarD^I27W
No CarD
CarD^K125E
CarD^WT

Figure 2


A
Normalized Equilibrium Fold Change over RNAP
1.0
0.8
0.6
0.5
0.4
0.2
0
10^0
10^1
10^2
10^3
10^4
Concentration of CarD (nM)
CarD^{WT}
CarD^{I27F}
CarD^{I27W}
B
k_{obs}^3 / s
0.03
0.02
0.01
0
0
100
200
300
400
500
Concentration of CarD (nM)

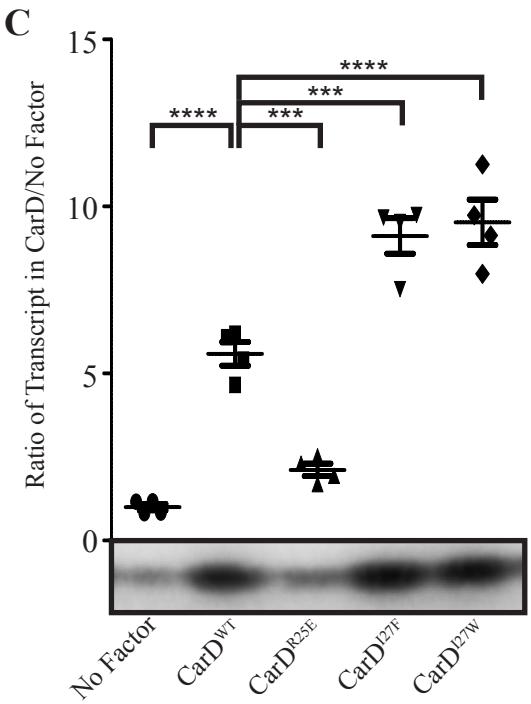

C
Ratio of Transcript in CarD/No Factor
15
10
5
0
****
***
***
****
No Factor
CarD^{WT}
CarD^{R25E}
CarD^{I27F}
CarD^{I27W}

Figure 3

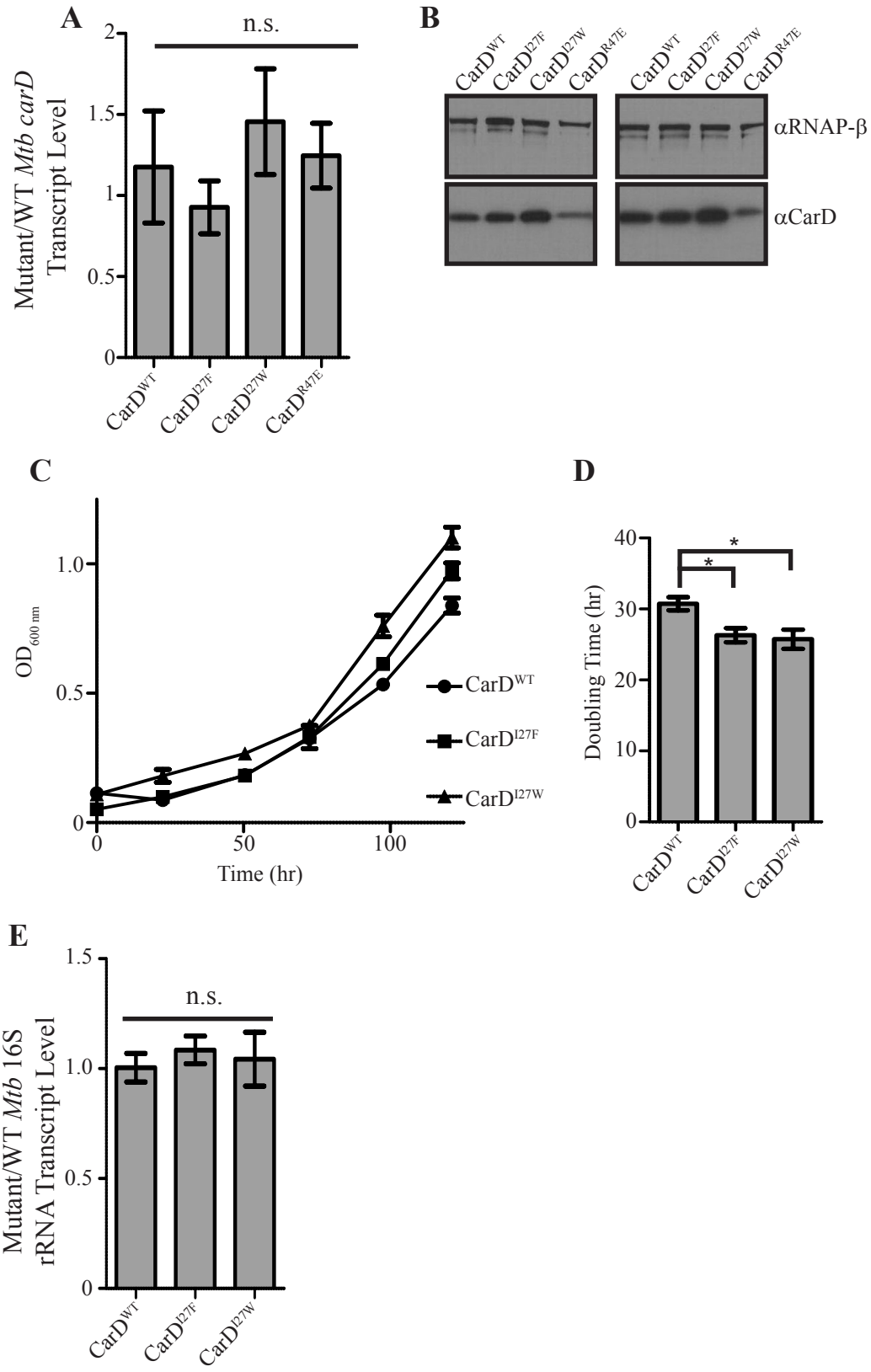
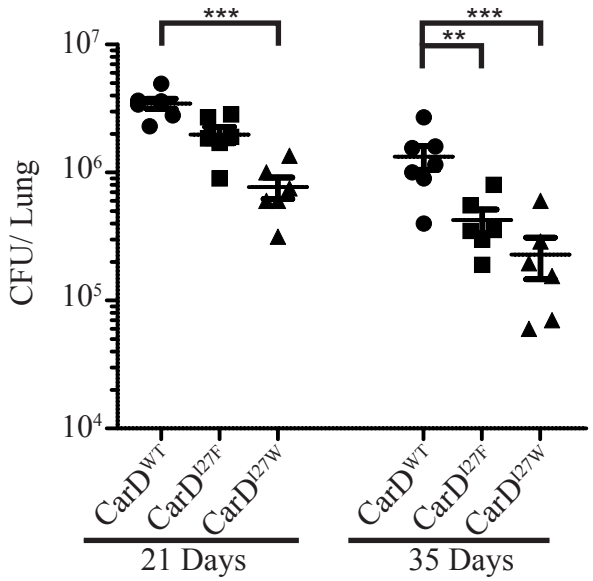

A
n.s.
Mutant/WT Mtb carD Transcript Level
2
1.5
1
0.5
0
CarDWT
CarDI27F
CarDI27W
CarDR47E
B
CarDWT
CarDI27F
CarDI27W
CarDR47E
CarDWT
CarDI27F
CarDI27W
CarDR47E
αRNAP-β
αCarD
C
OD600 nm
1.0
0.5
0
0
50
100
Time (hr)
CarDWT
CarDI27F
CarDI27W
D
Doubling Time (hr)
40
30
20
10
0
*
*
CarDWT
CarDI27F
CarDI27W
E
n.s.
Mutant/WT Mtb 16S rRNA Transcript Level
1.5
1.0
0.5
0
CarDWT
CarDI27F
CarDI27W

Figure 4


***
***
**
10^7
10^6
10^5
10^4
CFU/ Lung
$CarD^{WT}$
$CarD^{I27F}$
$CarD^{I27W}$
$CarD^{WT}$
$CarD^{I27F}$
$CarD^{I27W}$
21 Days
35 Days

Chapter 4: Defining the CarD regulon in Mycobacteria through RNA sequencing

Ashley Garner, Jahangheer Shaik, and Christina L. Stallings

A version of this chapter is being prepared for publication.

Abstract

CarD is an RNA polymerase (RNAP) binding protein conserved in numerous bacterial species and essential in mycobacteria. Chromatin immunoprecipitation sequencing reactions have shown that CarD is globally localized to promoters throughout the genome. However, thus far its function has only been studied in detail at ribosomal RNA (rRNA) promoters and consensus *Escherichia coli* promoters. In this study we utilize a panel of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* strains that express different concentrations of $CarD^{WT}$ or CarD mutants to define the *in vivo* CarD regulon in these species. The CarD mutations we examine individually target either CarD's interaction with RNAP or DNA. These studies are the first to investigate the regulatory activity of CarD at non-rRNA promoters and provide the first evidence that CarD is important for regulation of the majority of promoters in *M. tuberculosis*. We show that CarD activity is influenced by promoter sequence by identifying a correlation between promoter motifs and regulation by CarD. In *M. smegmatis*, we provide evidence of an effect of the nucleoid associated protein EspR on gene expression in CarD mutants. These studies provide new information regarding the global regulatory activity of CarD in mycobacteria and provide the first insights into determinants of CarD-mediated regulation.

Introduction

Ribosomal RNA (rRNA) operons include the most well characterized mycobacterial promoters, regulation of which is essential for growth and viability. Early studies of CarD found that depletion of CarD regulates rRNA and we therefore chose to focus our initial efforts to characterize CarD's activity as a transcriptional regulator on these operons. *In vivo*, CarD has been shown to activate transcription from rRNA promoters, and to be required for wild-type (WT) levels of rRNA in the cell (Chapter 2 Fig. 6, (87, 108)). Biochemically, we have shown

that CarD stabilizes the RNA polymerase (RNAP)-promoter open complex (RP_o) formed during transcription initiation (Chapter 2, Fig. 7, (87, 108, 109, 123)). These studies have contributed to a model in which CarD's N-terminal domain interacts with the RNAP β-subunit (Chapter 1 Fig. 1, (85, 101)). At transcription initiation complexes, this interaction with RNAP positions CarD's C-terminus to interact with DNA just upstream of the -10 promoter motif (Chapter 1 Fig. 3, (86, 87)). The interaction between CarD and DNA is electrostatic, mediated by a basic patch on CarD, and is sequence independent (Chapter 1 Fig. 3, (87, 108)). CarD interacts with the RP_o with high affinity and stabilizes RP_o by reducing the rate of collapse back to RNAP-promoter closed complex (RP_c). CarD interacts with RP_c with lower affinity, but at physiologically relevant concentrations (Chapter 2 Fig. 8, (109)). This interaction with RP_c promotes melting of the promoter and transition to RP_o (109). In addition to CarD's interactions with RNAP and with promoter DNA, CarD has a conserved tryptophan residue (W85) that has been proposed to wedge into the splayed minor groove at the upstream edge of the transcription bubble to stabilize RP_o (86, 87, 108). To date, this model of CarD activity has only been tested on rRNA promoters and consensus *Escherichia coli* promoters (86, 108, 109, 123).

A prominent open question in CarD research concerns its effect at non-rRNA promoters. Chromatin immunoprecipitation sequencing (ChIP-seq) experiments have shown that CarD is globally localized to promoters throughout the *Mycobacterium smegmatis* chromosome, suggesting it may be a general member of the transcription initiation machinery in mycobacteria (Chapter 1 Fig. 2, (87)). Additionally, several groups have shown that, compared to the *E. coli* RNAP, the mycobacterial RNAP is inherently deficient in formation of stable RP_o and may require CarD to form stable, initiation-competent complexes (109, 123). If CarD's activity at rRNA promoters is representative of its activity at all promoters, then we would expect a global

decrease in transcription in the absence of CarD. However, initial microarray experiments in *M. smegmatis* that compared CarD-depleted cells to cells expressing CarD found only a small portion of the genome to be deregulated greater than 2-fold. Importantly, interpretation of these results is confounded by the cell death caused by CarD-depletion, which also affects transcript abundance (85). It is therefore difficult to determine the extent of the CarD regulon from this experiment.

During our studies, we have constructed numerous strains of mycobacteria in which CarD's ability to regulate transcription is affected without causing cell death. Decreasing the affinity of CarD's interaction with either RNAP or DNA causes *in vivo* phenotypes including decreased ability of CarD to activate transcription of rRNA promoters, decreased growth rate, and increased stress sensitivity (Chapter 1 Fig. 1, Chapter 2 Fig. 2, 3, 5, and 6, (87, 101, 108). Biochemically, we have shown that mutations that weaken these interactions attenuate CarD's ability to stabilize the open complex, which is crucial for CarD-mediated transcriptional regulation (Chapter 2 Fig 7, (87, 108, 109, 123)). Together these results show that CarD's interactions with both RNAP and DNA are crucial for its transcriptional regulatory activity (101, 108, 109). We have also developed mutants ($CarD^{I27F}$ and $CarD^{I27W}$) with increased affinity for RNAP (Chapter 3). These mutations stimulate of production of a 3nt aborted transcript and increase the half-life of RNAP-promoter complexes as compared to $CarD^{WT}$ (Chapter 3 Fig. 2, (136)). *In vivo*, these strains have similar expression of rRNA as $CarD^{WT}$ encoding strains but grow faster than the $CarD^{WT}$ strain (Chapter 3 Fig. 3, (136)). Additionally, CarD's regulatory activity is concentration-dependent: *in vitro* assays have shown that both the amount of open complex at equilibrium and the kinetics of RP_o formation are influenced by CarD concentration (109). *In vivo*, mycobacteria upregulate *carD* in response to numerous stresses, which suggests

that varying the concentration of CarD might affect gene expression (85). These viable strains with altered CarD activity represent a powerful tool with which we can assay CarD's regulatory role at non-rRNA promoters.

To determine the role of CarD in regulation of non-rRNA promoters and identify factors that may modify CarD's activity, we performed RNA sequencing (RNA-seq) analysis of the transcriptomes of *Mtb* strains that express either mutants with individually attenuated CarD activities (e.g. DNA binding, RNAP binding) or that express different concentrations of the CarD protein. We show that CarD is important for regulation of the majority of transcripts in *Mtb*, but observe variable effects on regulation of specific transcripts, which suggests CarD's activity may be affected by promoter characteristics or other factors such as genomic architecture or nucleoid associated proteins. By analyzing DNA sequences upstream of transcription start sites of regulated transcriptions, we found a correlation of -10 promoter motifs with transcripts deregulated upon alterations in CarD. We find that transcripts lacking a recognizable -10 promoter motif are more sensitive to alterations in CarD than promoters with recognizable motifs. These studies represent a major advance in our understanding of CarD activity at non-ribosomal promoters and its role in promoting efficient gene expression in the cell.

Experimental Procedures

Bacterial strains and growth conditions. All *Mtb* strains were derived from the Erdman strain and were grown at 37°C in Sauton's broth media (0.5 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4$, 4 g 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). Gene switching was used to construct the mgm3080, csm230, csm231, csm195, and csm196 strains of *M. tuberculosis* (expressing $CarD^{WT}$, $CarD^{I27F}$, $CarD^{I27W}$, $CarD^{R47E}$, and $CarD^{K125A}$, respectively) used in this study. Construction of these strains

has been detailed in previous studies (85, 101, 108, 136).

All strains used in this manuscript encode only one *carD* allele.

RNA Preparation. RNA was prepared from 30 ml of log-phase *M. tuberculosis* Mgm3080 ($CarD^{WT}$), csm230 ($CarD^{I27F}$), csm231 ($CarD^{I27W}$), csm195 ($CarD^{R47E}$), and csm195 ($CarD^{K125A}$) in triplicate, as previously described (85). 1-5 µg of DNA-free RNA from each replicate was submitted to the Genome Technology Access Center (GTAC) at Washington University. Ribozero was used to deplete rRNA and Illumina sequencing was used to determine the transcriptome of each strain.

Construction of Venn Diagrams. Venn diagrams were constructed through the use of either https://www.stefanjol.nl/venny or Venny website (137). Some final images in this paper were further modified.

Alignment of Sequencing Reads. The short reads from the next-generation sequencing machine are aligned to the genome of interest using Tophat software. Tophat allows splitting the reads to align to the genome to address alternate splicing. Since our genomes of interest do not contain introns, the choice of aligners is not an issue. After the reads were aligned to the genome, we found the average coverage per gene using HTSEQ package (138). HTseq allows counting the reads per gene, per exon or any other genomic feature as indicated in the general feature format (GFF) file. The average coverage per gene depends on the two factors: the biology and the total number of reads sequenced per sample. In order to eliminate the bias introduced by different read counts in different samples, we employed "upper quantile" normalization. The genes that are differentially expressed between different experimental conditions are found using DEseq package (139). We applied a false discovery rate of 0.01 to select genes that are

differentially expressed between multiple experimental conditions. All the plots were generated using custom scripts in R.

Results

Decreasing CarD's affinity for RNAP or DNA causes non-uniform deregulation of expression of the majority of the genes in *Mtb* with a high degree of similarity between the two sets of deregulated genes. In previous work, we constructed strains of *Mtb* in which a copy of *carD* is expressed from a constitutive promoter at the *attB* site, a non-endogenous chromosomal locus, and the endogenous copy of *carD* is deleted (85). We have previously shown that the decreasing CarD's affinity for RNAP ($CarD^{R47E}$) or DNA ($CarD^{K125A}$) causes decreased growth rate, deregulated rRNA expression, and increased stress sensitivity (Chapter 1 Fig. 1, Chapter 2 Fig. 2, 3, 5, and 6, (87, 101, 108)). While the $CarD^{K125A}$ mutation does not affect CarD's ability to bind DNA in an EMSA assay, the mutation is within the DNA binding basic patch and has phenotypes consistent with an effect on DNA binding *in vivo* (Chapter 2, (108)). We have shown biochemically that weakening the interaction with RNAP or DNA decreases CarD's ability to stabilize open complex at rRNA promoters (Chapter 2 Fig. 7, (108, 109)). To determine the effect of these mutations on transcriptional regulation of non-rRNA promoters, we compared the transcriptomes of the strains expressing $CarD^{R47E}$ or $CarD^{K125A}$ to a strain expressing $CarD^{WT}$ from the same chromosomal locus (Table 1), (mgm3080, designated as $CarD^{WT}$). Importantly, while the *carD* transcript is expressed similarly, we have previously shown that the $CarD^{R47E}$ strain has a lower cellular concentration of CarD than the $CarD^{WT}$ strain (Chapter 3, Fig. 3). Decreasing CarD's affinity for RNAP had global effects on the *Mtb* transcriptome, significantly deregulating 2,454 transcripts, or 61% of the transcriptome when a false discovery rate (FDR) of less than 0.01 is used to determine significance and no fold-change

cut off is imposed (Fig. 1A). The fold change of deregulated transcripts varies from 1.23-8.36 fold-upregulated and 1.24-3.93 fold-downregulated with an average of 1.57-fold upregulated or 1.52 fold downregulated. If we use this average deregulation as a fold change cut off and only look at transcripts deregulated greater than 1.5-fold, $CarD^{R47E}$ upregulated 611 transcripts and downregulates 483. The $CarD^{K125A}$ mutation also had dramatic effects on the transcriptome, significantly deregulating 2,247 transcripts, of which 743 were deregulated greater than 1.5-fold (Fig. 1B). In this strain, transcripts were significantly upregulated between 1.12 and 43.11 (average of 1.53) while transcripts were downregulated between 1.12 and 6.10 (average of 1.42). Both the $CarD^{R47E}$ and $CarD^{K125A}$ mutants decrease growth rate, decrease rRNA levels, increase antibiotic sensitivity, decrease virulence, and attenuate CarD's ability to stabilize RP_o. The phenotypic similarity of these two mutants is reflected in their effect on the transcriptome of *Mtb*, as 86.1% of transcripts significantly down regulated by $CarD^{K125A}$ and 80.4% of transcripts upregulated by $CarD^{K125A}$ without a fold change cut off, are similarly down or upregulated by $CarD^{R47E}$ (Fig 1C). Only 0.3% of transcripts downregulated by $CarD^{K125A}$ and 1.2% of transcripts upregulated by $CarD^{K125A}$ are inversely regulated in the $CarD^{R47E}$ strain. This indicates that CarD's interactions with RNAP and/ or DNA are broadly required for CarD mediated gene regulation and supports phenotypic analysis that suggested that the requirement for these activities are linked such that transcripts are more likely to require both than either independently.

The one phenotype that is not shared between mutants with decreased affinity for RNAP or DNA is sensitivity to oxidative stress, which is observed only in mutants with decreased affinity for RNAP or mutation of the conserved W85 residue (Chapter 1 Fig. 1, Chapter 2 Fig. 3, (101, 108)). Based on this difference in sensitivity to oxidative stress, it was suggested that there

are transcripts deregulated by decreasing CarD's affinity for RNAP that are not deregulated by decreasing CarD's affinity for DNA (Chapter 2, (108)). While the CarD^{R47E} strain does upregulate 301 transcripts and downregulate 264 transcripts that are not shared with CarD^{K125A} and CarD^{K125A} upregulates 232 and downregulates 135 transcripts that are not shared with CarD^{R47E}, we found no differences in regulation of examined genes known to respond to, or regulate the response to, oxidative stress. Importantly, this study is examining the transcriptome of exponential cultures not being subjected to oxidative stress. It remains likely that these two mutants have distinct transcriptional responses to oxidative stress that simply isn't apparent in the absence of the stress.

Analysis of the transcriptome of loss-of-function CarD mutants suggests that CarD^{WT} can act as a transcriptional repressor. At rRNA promoters, decreasing CarD's affinity for RNAP or DNA has been shown to decrease CarD's effect on stabilization of RP_o *in vitro*, which decreases transcription from these promoters *in vivo* (Chapter 2 Fig. 6 and 7, (87, 108, 109)). If CarD similarly activates transcription from all promoters it regulates, then decreasing CarD's affinity for either RNAP or DNA would uniformly result in downregulation of directly regulated transcripts. This is not supported by the RNA-seq data as transcripts deregulated in either CarD^{R47E} or CarD^{K125A} are slightly more likely to be upregulated in these mutants than downregulated (53.8% and 56.2% of transcripts deregulated in CarD^{R47E} and CarD^{K125A}, respectively, are upregulated, Fig. 1AB). While upregulation of some transcripts may result from indirect effects through deregulation of other transcription regulators, the prevalence of upregulated transcripts in these strains supports a direct role for CarD in the repression of transcripts. Assuming that CarD stabilizes RP_o at all regulated promoters, CarD likely represses transcription by hyperstabilizing RP_o and inhibiting promoter escape or transition to the

elongation complex. The ability of a transcription factor that regulates transcription by stabilizing an RNAP-promoter complex to activate or repress transcription in a promoter-specific manner has previously been described for the phage ϕ29 protein p4 (124). While the ability of CarD to repress transcription must still be demonstrated biochemically, these data are the first to suggest that this may be a physiologically relevant activity of CarD.

Increasing CarD's affinity for RNAP deregulates the transcriptome differently than mutations than loss of function mutants. We have previously characterized two mutants of isoleucine 27 in CarD's N-terminal domain ($CarD^{I27F}$ and $CarD^{I27W}$) that both increase CarD's affinity for RNAP (Chapter 3, (136)). *In vivo*, increasing CarD's affinity for RNAP causes less severe phenotypes than decreasing CarD's affinity for RNAP or DNA: $CarD^{I27F}$ and $CarD^{I27W}$ are not sensitized to any stress that has been tested (treatment with streptomycin, rifampicin, ciprofloxacin, ethambutol, or hydrogen peroxide), nor do these mutations cause deregulation of rRNA (Chapter 3 Fig. 3). Increasing CarD's affinity for RNAP results in faster growth than the $CarD^{WT}$ strain in liquid *Mtb* cultures, however it does not increase the fitness of *Mtb* as these mutations cause a virulence defect in mice (Chapter 3 Fig. 3 and 4, (136)). In keeping with these mild *in vivo* phenotypes, when we compare the transcriptomes of the high affinity mutants to the $CarD^{WT}$ strain we find that increasing CarD's affinity for RNAP has a smaller effect on the transcriptome than was observed in the loss-of-function mutants. In the $CarD^{I27F}$ and $CarD^{I27W}$ strains, a total of 912 and 1,333 transcripts are deregulated, respectively, compared to the $CarD^{WT}$ strain when a FDR of less than 0.01 is used to determine significance and no fold change cut off is imposed on the data (Fig. 2A). The fold change of up and downregulated transcripts, respectively, in the $CarD^{I27F}$ strain varies from 1.23-8.36-fold change with an average of 1.56 and 1.24-3.93 fold change with an average of 1.53. Similarly, the fold change of

transcripts in the CarD^{I27W} strain varies from 1.22-13.96-fold change with an average of 1.57 for upregulated transcripts and 1.21-4.20 with an average of 1.53 for downregulated transcripts. When the data are constrained by a 1.5-fold change cut off, 390 (250 upregulated, 140 downregulated) and 590 (365 upregulated, 224 downregulated) transcripts are deregulated in CarD^{I27F} or CarD^{I27W}, respectively. The I27F and I27W mutations both strengthen the affinity of CarD for RNAP, but this effect is more pronounced in the I27W mutant (Chapter 3 Fig. 1, (136)). Similarly, the transcripts deregulated in the CarD^{I27F} and CarD^{I27W} strains overlap heavily, but a larger number of transcripts are deregulated in the CarD^{I27W} strain (Fig. 2C). There are no transcripts upregulated in one high-affinity mutant and downregulated in the other, which underlines the similarity in the effects of these two mutations.

CarD has a stronger affinity for RNAP and is more highly expressed in the CarD^{I27F} and CarD^{I27W} strains than the CarD^{WT} strains; conversely, CarD has a weakened affinity for RNAP and is less highly expressed in the CarD^{R47E} strain than the CarD^{WT} strain. We would therefore predict that these two classes of CarD mutants should have opposite effects on the transcriptome. We find that 69.8% and 68.3% of transcripts deregulated in CarD^{I27F} and CarD^{I27W}, respectively, are also deregulated in CarD^{R47E} (Fig. 3AB). Importantly, 30% of transcripts deregulated in these mutants were not deregulated in the CarD^{R47E} strain, which indicates that transcripts deregulated in individual mutants do not completely define the CarD regulon and highlights the importance of testing the effect of numerous alterations in CarD on transcript regulation. CarD^{I27F} and CarD^{I27W}, respectively, have the opposite effect as CarD^{R47E} on regulation of 72.4% and 68.9% of the transcripts that are deregulated in both CarD^{R47E} and the indicated I27 mutants, which supports our prediction based on our understanding of these CarD mutants. We predict that the

transcripts that are similarly deregulated in the I27 mutants and the CarD^{R47E} mutant are likely the result of indirect regulation through deregulation of other transcription factors.

Increased concentrations of CarD^{WT} are correlated with increased growth rate in *Mtb*. We have previously demonstrated different concentrations of CarD in strains of *Mtb* expressing CarD mutants with altered affinity for RNAP (Chapter 3, Fig.3). To directly assay the effect of CarD concentration on gene expression we wanted to compare two strains expressing the same allele of CarD but at different cellular concentrations. When we compare expression of the *carD* allele from the endogenous locus to expression of *carD* from the non-endogenous locus used in our engineered CarD strains we find that *carD* is expressed 3.4 fold higher in the WT *Mtb* strain than the CarD^{WT} strain (Fig. 4A). This difference in gene expression is reflected in protein concentration as the WT *Mtb* strain also expressed higher concentration of the CarD protein (Fig. 4B). This higher cellular concentration of CarD has phenotypic effects on the bacterium including increased growth rate (Fig. 4C) in the WT *Mtb* strain. Interestingly, the growth rate in WT *Mtb* is comparable to the growth rate previously reported in the CarD mutants with higher affinity (24.2, 26.3, and 25.8 hours doubling time for WT *Mtb*, CarD^{I27F}, and CarD^{I27W} respectively) despite the WT *Mtb* strain having higher cellular concentrations of CarD than the high affinity mutants (Fig. 4B). Therefore, increasing the affinity of CarD to the RNAP can compensate for the negative effect of lower levels of *carD* expression on growth rate. These results, as well as the growth defect previously reported in a strain expressing a CarD mutant with lower affinity for RNAP (101), suggest that a minimum amount of CarD must associate with RNAP at initiation complexes for WT-like growth. Importantly, we have previously shown similar expression of rRNA in the high affinity mutants and the CarD^{WT} strain, suggesting that CarD's effect on growth can be independent from its effect on rRNA regulation (Chapter 3, Fig.

3). To determine in the higher concentration of CarD in the WT Mtb strain was rescuing growth through higher expression of rRNA we performed qRT-PCR on exponential cultures of WT *Mtb* and the CarD^{WT} strain. Similar to results from the high affinity mutants, we found similar expression of rRNA in these strains (Fig. 4D).

Increasing the concentration of CarD for RNAP attenuates survival of *Mtb* during infection of mice. To determine if the concentration of CarD affects virulence, we compared the virulence of WT *Mtb* and CarD^{WT} in C57BL/6J mice in the lungs at 21 and 35 dpi (Fig. 4E). We found that at 35 dpi WT *Mtb* had significantly lower bacterial burdens in the lungs than the CarD^{WT} strain (Fig. 4E), which expresses lower levels of *carD* transcript and has lower cellular concentrations of CarD protein (Fig. 4AB). Therefore, the CarD concentration-dependent growth defect of the CarD^{WT} strain does not cause a virulence defect in mice and in fact correlates with higher bacterial burden in early chronic infection. These data highlight the impact of differences in CarD protein concentration on the fitness of *Mtb* and may support the importance of regulating CarD protein levels.

Changes in CarD concentration cause significant, non-uniform deregulation of the majority of the transcriptome. The differences in both growth rate and virulence of the CarD^{WT} and the WT *Mtb* strains indicate that the cellular concentration of CarD *in vivo* can affect gene expression. This corresponds to *in vitro* data in which, when experiments are performed with a constant concentration of RNAP, the amount of RP_o formed at equilibrium is dependent on the concentration of CarD in the reaction (109). To determine how altering CarD's cellular concentration affects CarD's regulon, we performed RNA-seq on exponential cultures to compare the WT *Mtb*, which expresses *carD* from the endogenous locus, and the CarD^{WT} strain, in which *carD* is expressed at a lower level from the *attB* locus. To be consistent with other

comparisons in this study, we designated the CarD^{WT} strain as the control. When we use a FDR cut off of less than 0.01 to determine significance and do not impose a fold change cut off, 52% of the transcriptome is significantly deregulated in the WT *Mtb* strain when compared to the CarD^{WT} strain (Fig. 5A). Within these deregulated transcripts, we find a similar number of transcripts up- and downregulated, with large variations in the fold change in expression. Downregulated transcripts vary from 1.11 to 60.24 fold lower in WT *Mtb*, with an average fold change of 1.39. Upregulated transcripts vary from 1.15 to 8.25 fold higher in the WT *Mtb* strain, with an average fold change of 1.50. This variability in response to altered CarD concentration suggests that CarD's activity may be influenced by promoter specificity or other intrinsic properties of these transcripts, such as genomic location, in a manner that is affect by concentration of CarD as well as alterations in its activities. If we impose a fold-change cut-off on the data and only look at deregulated transcripts with greater than 1.5-fold changes, then 315 transcripts are downregulated in WT *Mtb* and 283 transcripts are upregulated, for a total of 598 transcripts or 15% of the transcriptome. The dramatic decrease in the number of genes affected by alteration in CarD concentration upon implementation of a fold-change cut off indicates that changes in CarD concentration has a subtle but significant effect on the majority of transcripts that are sensitive to CarD concentration. Importantly, there is a sizeable portion of the transcriptome that is more substantially affected by CarD concentration and it is likely deregulation of these transcripts that causes the growth defect in the CarD^{WT} strain (Chapter 3 Fig. 3, (136)).

Both concentration and affinity are important drivers of transcript regulation in CarD mutants with increased affinity for RNAP. In addition to their direct effects on RNAP affinity, mutations of CarD's RNAP interacting domain (RID) also affect the cellular

concentration of CarD. Specifically, we have shown that the cellular concentration of CarD protein in strains that express *carD* from the *attB* site is proportional to CarD's affinity for RNAP with CarD^{R47E} having the lowest affinity and concentration, followed by CarD^{WT}, CarD^{I27F}, and CarD^{I27W}, which has both the highest affinity and the highest cellular concentration. The WT *Mtb* strain, which has higher transcriptional expression of *carD*, has the highest cellular concentration of CarD protein (Chapter 3 Fig. 3, (136)). Our RNA-seq analysis of the WT *Mtb* strain demonstrated that gene expression can be affected by CarD protein concentration. Both the similarity of the CarD^{R47E} and CarD^{K125A} transcriptomes and our *in vitro* analysis of CarD mutants indicate that affecting CarD's affinity for RNAP directly affect gene expression in a concentration-independent manner. To gain insight into the roles for alterations in concentration and affinity in the deregulation of transcripts in the CarD RID mutants we analyzed the expression of individual transcripts within each mutant in comparison to either WT *Mtb* or CarD^{WT} as well as in comparison of the two CarD^{WT} encoding strains to each other.

We initially examined the roles for affinity and concentration in deregulation of transcripts in the high affinity CarD mutants. Because we know both the relative CarD concentrations and affinities of CarD for RNAP in these four strains (CarD^{WT}, WT *Mtb*, CarD^{I27F}, and CarD^{I27W}), one can predict gene expression patterns for various regulatory possibilities (Table 2). Since both CarD^{I27F} and CarD^{I27W} have higher affinity for RNAP than either strain that encodes a WT allele of CarD and a cellular concentration of CarD that falls between the two WT encoding strains we would predict similar gene expression profiles in response to different regulatory possibilities and will therefore refer to the two mutants collectively during this discussion.

In the simplest regulatory schemes, regulation of a gene may be exclusively affected by either alteration of affinity (scheme A) or concentration (scheme B). If regulation of a gene is exclusively affected by alterations of affinity (scheme A), we would expect the gene to be either up or down regulated in the I27 mutant compared to both strains encoding a WT allele of CarD. Since the two strains encoding a WT allele of CarD have the same affinity for RNAP, we would expect the gene not be deregulated when they are compared to each other. If we examine the genes significantly deregulated in $CarD^{I27F}$ without a fold-change cut off, we find that 135 fit this expression profile with 44 downregulated in $CarD^{I27F}$ compared to either of the WT encoding strains and 91 upregulated. In the $CarD^{I27W}$ strain, 89 downregulated and 136 upregulated transcripts fit this expression profile. Transcripts that fit this expression profile are the only ones deregulated in the I27 mutants that solely respond to differences in affinity but this regulatory scheme only account for 10% of genes deregulated in $CarD^{I27F}$ and 12% of genes deregulated in $CarD^{I27W}$, which indicates that differences in concentration are an important driver of deregulation in these strains.

Alternatively, if deregulation of a transcript is exclusively caused by alterations in concentration of CarD (scheme B), we would predict that a transcript would be up or down regulated in the comparison of $CarD^{WT}$ to the WT *Mtb* strain, dependent on whether more CarD increases or decreases transcription of that transcript. Because the I27 mutants have an intermediate concentration, we would expect they would have the same effect on gene expression as the WT Mtb strain when compared to the $CarD^{WT}$ strain but the opposite effect when compared to the WT *Mtb* strain (e.g. if a transcript is upregulated in WT *Mtb* compared to $CarD^{WT}$, we would expect it would be upregulated in an I27 mutant compared to $CarD^{WT}$ but downregulated in the I27 mutant compared to WT *Mtb*). Importantly, we would predict to see

this same expression profile if changes in concentration are the primary driver of deregulated but affinity had a minor effect (scheme D1), so we cannot totally rule out an effect of affinity in regulation of these transcripts. In the CarD^{I27F} mutant, we find 43 transcripts with this expression profile, of which 29 transcripts have lower expression in WT *Mtb* compared to CarD^{WT} and 14 have higher expression in the WT *Mtb* strain. In the CarD^{I27W} mutant, 35 transcripts fit this expression profile and have lower expression in the WT *Mtb* strain and 17 fit this expression profile and have higher expression in the WT *Mtb* strain when it is compared to the CarD^{WT} strain. As only 3% of the transcripts deregulated in either mutant fit this expression profile, we can conclude that the majority of the transcripts deregulated in the I27 mutants are likely due to a combination of affinity and concentration.

If a gene is affected by both concentration and affinity (scheme C), then it could be regulated in numerous ways. The first regulation scheme we considered is if were equally affected both concentration and affinity and if both factors affected regulation in a similar manner such that both higher affinity and higher concentration led to either more or less expression of the transcript (Scheme C1). In the case that both concentration and affinity increase expression of the transcript, then we would expect the transcript to be upregulated in the I27 mutant compared to CarD^{WT} because the mutant has higher affinity. When the I27 mutant is compared to WT *Mtb*, the mutant would have higher affinity but WT *Mtb* would have higher concentration. In this scheme, we predict these two effects would balance each other and we would see no change in expression. WT *Mtb* has higher CarD concentration than CarD^{WT}, so we would expect the transcript to also be upregulated in WT *Mtb* relative to CarD^{WT}. 373 transcripts deregulated in CarD^{I27F} and 459 transcripts deregulated in CarD^{I27W} have this expression trend. If increasing both concentration and expression led to less transcript expression, we would expect

the gene to be down in the I27 mutants and WT *Mtb* relative to CarD^{WT} and equal in the I27 mutants relative to WT *Mtb*. We see this expression pattern to 220 transcripts deregulated in CarD^{I27F} and 323 transcripts in CarD^{I27W}. We also considered the regulatory scheme in which concentration and affinity are equally important but have opposite effects on transcript expression (scheme C2). If increasing concentration and decreasing CarD's affinity for RNAP increases transcript expression, then we would expect the transcript to be similarly expressed in the CarD^{WT} strain and the I27 mutant since the mutant has both higher expression and higher affinity. The transcript would be down regulated in the I27 mutant relative to WT *Mtb* as WT *Mtb* has higher expression and lower affinity of CarD for RNAP than the I27 mutants. The transcript would be upregulated in WT *Mtb* compared to the CarD^{WT} strain due to the higher CarD concentration in the WT *Mtb* strain. This matches the expression data for 196 transcripts deregulated in CarD^{I27F} and 212 transcripts deregulated in CarD^{I27W}. If decreasing concentration and increasing affinity increases transcript expression, then we would expect no difference in the I27 mutants relative to CarD^{WT} but upregulation relative to the WT *Mtb* strain. The transcript would be downregulated in WT Mtb relative to CarD^{WT}. This expression scheme matches the expression profile of 201 transcripts deregulated in CarD^{I27F} and 180 transcripts deregulated in CarD^{I27W}. Of all the gene expression schemes we considered, these two schemes in which concentration and affinity are equally important explain the regulatory pattern of the largest number of genes, together accounting for regulation of 71% of transcripts deregulated in CarD^{I27F} and 64% of transcripts deregulated in CarD^{I27W}. The large numbers of transcripts explained by regulatory schemes in which concentration and affinity are equally important indicate that both are key drivers of regulation of CarD activity.

We next considered regulatory schemes in which contribution of concentration and affinity are not equal. If we considered concentration to be primary driver of deregulation (scheme D1) we predicted the same expression profiles as when concentration was the only cause of deregulation (scheme B). However, if affinity is dominant to concentration and both have the same effect on regulation, such that increasing concentration and affinity increases gene expression, we predict a gene expression profile that is unique to this analysis (scheme D2). The transcript would be upregulated in the I27 mutants relative to either strain encoding a WT allele of CarD because of the higher affinity of the mutants for RNAP. The transcript would also be upregulated in the WT *Mtb* strain relative to the $CarD^{WT}$ strain because of the higher concentration of CarD in the WT *Mtb* strain. 24 transcripts in $CarD^{I27F}$ and 73 transcripts in $CarD^{I27W}$ have this expression profile. Concentration and affinity may also be negatively correlated with expression such that increasing either decreases transcript expression. Here we would expect down regulation the in the I27 mutants relative to either strain expressing a WT allele of CarD and down regulation of the transcript in WT *Mtb* relative to $CarD^{WT}$. We observe this expression profile for 17 transcripts deregulated in $CarD^{I27F}$ and 42 transcripts deregulated in $CarD^{I27W}$. We also considered a scheme in which affinity is the dominant driver of affinity but regulation is also affected by concentration but it has the opposite effect (scheme D3). If higher affinity and lower concentration increase expression, then a transcript would be upregulated in the mutant relative to either strain expressing a WT allele of CarD because of affinity but the same transcript would be downregulated in WT *Mtb* relative to the $CarD^{WT}$ strain because of the increased concentration of CarD in WT *Mtb*. We find 25 transcripts deregulated in $CarD^{I27F}$ and 32 transcripts deregulated in $CarD^{I27W}$ with this expression pattern. If higher affinity and lower concentration decrease expression, then the transcript should be downregulated in the I27

mutants and upregulated in WT *Mtb* when compared to the CarD^{WT} strain. This expression pattern is observed for 10 transcripts deregulated in CarD^{I27F} and 14 transcripts deregulated in CarD^{I27W}. While a substantially smaller number of deregulated transcripts follow this scheme than when we considered concentration and affinity to have equal effects, we find that affinity if more than 5 times as likely to be the dominant driver of deregulation than concentration in both mutants (the number of transcripts in schemes A, D2, and D3 versus the number of transcripts in schemes B/D1) underscoring the role for increasing CarD's affinity for RNAP in directly affecting transcript expression.

In total, we have been able to dissect the roles of affinity and concentration in the regulation of 89.3% of transcripts deregulated in CarD^{I27F} and 87.9% of transcripts deregulated in CarD^{I27W}. The transcripts deregulated in CarD^{I27F} and CarD^{I27W} are highly similar which would predict that many of the transcripts deregulated in CarD^{I27F} should also be deregulated in CarD^{I27W} and that they should be in the same regulatory scheme in each mutant. The data supports this with 81.6% of transcripts deregulated in CarD^{I27F} with identified regulatory schemes are also deregulated with the same regulatory scheme in CarD^{I27W}.

It is more difficult to separate the effect of decreased CarD concentration from the effect of decreased RNAP affinity in the CarD^{R47E} strain, as the concentration of CarD in the CarD^{R47E} strain is lower than the concentration of CarD in either the CarD^{WT} strain or WT *Mtb* (136). This results in many of the regulatory schemes discussed above yielding the same gene expression profile. Because this limits our ability to interpret the data we did not do this type of analysis for CarD^{R47E}.

Transcripts lacking a recognizable -10 promoter motif are more sensitive to alterations in CarD. Transcriptional expression of transcripts is usually regulated, at least in

part, by the promoter that drives its expression. Structural studies have mapped CarD's interaction with DNA to the bases from -10 to -14 relative to the TSS (86), which partially overlaps with the -10 promoter motif, which suggests that CarD's regulatory activity may be affected by promoter sequence. It has been very difficult to identify promoters in mycobacteria in part because of the GC richness of their genome, which has made it difficult to correlate promoter sequences to RNA-seq data. Identification of promoters in *Mtb* was recently facilitated by identification of the primary transcription start sites (TSS) of 1,778 transcripts. The sequences upstream of the TSSs were mined for promoter motifs (140). There are 4,037 identified transcripts in *Mtb* which includes genes, sRNA, and tRNAs. As many mycobacterial genes are in operons, these 1,778 primary TSSs likely account for the majority of TSSs in the genome. A consensus mycobacterial -10 motif of TANNNT was detectable in the first 20 bp upstream of the TSS for 72.3% of these transcripts. Of the transcripts that lack a TANNNT motif, a GNNANNNT motif was detected for just over half of the remaining TSSs, accounting for 15.6% of all the transcripts with a mapped TSS. For the remaining 12.1% of mapped TSSs, there was no detectable -10 motif. No -35 motif was detected in first 50 bp upstream for any of the mapped TSSs (140).

To determine if the -10 motif is predictive of deregulation upon mutation or alteration in concentrations of CarD, we compiled a list of transcripts with mapped TSSs that were up and/or downregulated greater than 1.5-fold in at least one of the WT *Mtb*, CarD^{I27F}, CarD^{I27W}, CarD^{R47E}, or CarD^{K125A} strains when compared to the CarD^{WT} strain. Transcripts that were deregulated in multiple mutants are only counted once in this analysis, for a total of 750 promoter sequences. As we have throughout this study, we used the CarD^{WT} strain as a reference strain. If we calculate the frequency of different -10 motifs of the deregulated transcripts we find that the

frequency of a TANNNT motif has decreased from 72.3% to 70.0% and the frequency of the transcripts without a detectable -10 motif has increased from 12.1% to 14.9%, compared to the frequencies of these motifs in TSSs throughout the genome (Fig 6A). The frequency of the GNNANNNT motif did not change (15.1% vs. 15.6% in all genomic TSSs). This trend of an increased frequency of promoters that lack a -10 motif in transcripts deregulated by alterations in CarD is more pronounced as the stringency of the fold-change cut off is increased from 1.5 to 3-fold change. When TSSs for deregulated transcripts in the $CarD^{WT}$ strain are examined with either a 2- or 3-fold cut off, the distribution of TANNNT, GNNANNNT, and unrecognizable promoter motifs in this sample is statistically significantly different than the distribution of promoter motifs in the genome as a whole (Fig. 6A). While only 57 transcripts deregulated greater than 3-fold in at least one mutant have a mapped TSS, the correlation between lacking a recognizable promoter motif and deregulation is most apparent at this cut off, accounting for 28.1% of all promoters, which is more than twice the rate of this promoter type in the genome as a whole. This indicates that the promoters lacking a recognizable motif are both more likely to be deregulated by alterations in CarD and the promoters are more likely to be deregulated by a larger fold change than transcripts driven by promoters with different -10 motifs.

To determine if promoters lacking a recognizable motif were particularly correlated with a given alteration in CarD, we separated the 750 transcripts with a mapped TSS that were deregulated greater than 1.5-fold in a least one mutant by strain and direction of deregulation in comparison to the $CarD^{WT}$ strain. In this analysis, any promoter that is deregulated in more than strain is included in each analysis to which it is relevant; therefore, many promoter motifs are considered in multiple comparisons. In every strain analyzed (WT *Mtb*, $CarD^{R47E}$, $CarD^{K125A}$, $CarD^{I27F}$, and $CarD^{I27W}$), the distribution of promoter motifs associated with upregulated

transcripts is significantly different than the genome average. Upregulated transcripts are more likely to be associated with promoters lacking a motif and less often associated with TANNNT motif promoters than transcripts that were not upregulated by alterations in CarD (Fig. 6B). In addition to being less likely to be deregulated greater than 1.5 fold by alterations of CarD in any strain, promoters that contain a TANNNT motif are also likely to be deregulated in a smaller number of strains. Specifically, within the promoters deregulated greater than 1.5-fold, TANNNT motif promoters are deregulated in an average of 1.9 strains while promoters without a recognizable motif are deregulated in an average of 2.4 strains (Fig. 6C). The relatively high prevalence of promoters without a recognizable -10 motif driving expression of deregulated transcripts suggests that these promoters are particularly sensitive to alterations in CarD. Both CarD and -10 promoter motifs are known to contribute to formation of RP_o, which may explain the sensitivity of promoters lacking a recognizable -10 motif to alterations in CarD. Alternatively, the promoter without a recognizable -10 motif may be promoters recognized by alternative sigma factors. Nothing is known about CarD mediated regulation of holo-RNAP associated with alternative sigma factors, however several ways that these transcripts may be affected by alterations in CarD could be proposed. Firstly, CarD may directly regulate transcription from promoters recognized by alternative sigma factors in such a way that makes them sensitive to alterations in CarD. Secondly, alternative sigma factors may not associate with CarD and the apparent upregulation of these transcripts may reflect global changes in housekeeping sigma factor driven transcripts resulting in an apparent upregulation of alternative sigma factor transcripts. Thirdly, alterations in CarD may deregulate transcription in a way that decreases the bacterium's fitness and activates stress response pathways that are regulated by alternative sigma factors, resulting in an increase in the abundance of these transcripts as an

indirect effect of alterations in CarD. Importantly, all 3 promoter motifs are found associated with both CarD-sensitive and insensitive transcripts, which suggests that -10 promoter motifs do not completely explain CarD-mediated regulation (Fig. 6AB).

Effect of decreasing CarD's affinity for RNAP or DNA in *M. smegmatis.*

M. smegmatis is commonly used as a model organism to study *Mtb*, including studies of CarD. To test if CarD had the same effects on global gene expression in *M. smegmatis* as we observed in *Mtb,* we performed an RNA-seq experiment on strains of *M. smegmatis* expressing $CarD^{WT}$, a strain with decreased affinity for RNAP ($CarD^{R25E}$), and a strain with decreased affinity for DNA ($CarD^{K90A}$) from the *attB* site. *M. smegmatis* is less sensitive to alterations in CarD than *Mtb*, which allows for viability of more severely attenuated CarD mutants. To see the largest possible effect of alterations in CarD we used the $CarD^{R25E}$ mutant that has a more dramatic decrease in affinity for RNAP than $CarD^{R47E}$ (101) and the $CarD^{K90A}$ mutant that doesn't bind DNA in the EMSA assay (108) (Table 3). Importantly, in *M. smegmatis*, decreasing CarD's affinity for RNAP does not decrease the cellular concentration of CarD so all 3 strains have similar concentrations of CarD (Fig. 7A).

When the transcriptome of the *M. smegmatis* $CarD^{R25E}$ strain is compared to the $CarD^{WT}$ strain and a FDR of less than 0.01 is used to determine significance only 95 transcripts or 1% of the transcriptome is significantly deregulated. Of these transcripts 76 are downregulated and 19 are upregulated (Fig. 7B). With a FDR cut off of less than 0.01, the smallest fold change to be considered significant is 2.16-fold change from $CarD^{WT}$, likely because the $CarD^{WT}$ replicates in this experiment were somewhat variable than in the *Mtb* data set. Results of this comparison are very different than when the $CarD^{R47E}$ mutant is compared to the $CarD^{WT}$ strain in *Mtb*. One of the most apparent differences is that a much smaller percent of the genome is deregulated in the

M. smegmatis comparison. If the *Mtb* CarD^{R47E} data set is analyzed with a 2.16-fold cut off, 227 transcripts or 6% of the *Mtb* transcriptome are still deregulated, suggesting the higher variability of the *M. smegmatis* does not completely explain the differences in the transcriptome. Of the 227 transcripts deregulated greater than 2.16-fold in the *Mtb* CarD^{R47E}, 34% are downregulated, which is substantially different than the *M. smegmatis* CarD^{R25E} strain where 80% of transcripts are downregulated. The smaller percent of the genome deregulated by decreasing CarD's affinity for RNAP in *M. smegmatis* that *Mtb* may indicate differences in the regulatory role of CarD in the two organisms or that the larger percentage of the genome deregulated in *Mtb* is a result of the decreased concentration of CarD in the CarD^{R47E} strain compared to the CarD^{WT} strain in *Mtb*.

When the transcriptome of the *M. smegmatis* CarD^{K90A} strain is compared to the CarD^{WT} strain and a FDR of less than 0.01 is used to determine significance, 229 transcripts or 3% of the genome is deregulated (Fig. 7C). The smallest fold change that is significant in this data set is 1.92-fold different than WT. If this same fold change cut off is applied the to the comparison between CarD^{K125A} and CarD^{WT} in *Mtb*, then 6% of the genome is deregulated, indicating that a smaller percentage of the genome is deregulated in *M. smegmatis* for this mutant as well. Only 30% of deregulated transcripts are upregulated in the CarD^{K90A} strain, which is a larger percentage than was upregulated in the CarD^{R25E} strain but still less than the 72% of transcripts deregulated greater than 1.9-fold that were upregulated in the CarD^{K125A} comparison in *Mtb*. The sparsity of upregulated transcripts in the CarD^{R25E} and CarD^{K90A} strains may indicate that CarD^{WT} is less likely to repress transcription in *M. smegmatis* than *Mtb*. Both CarD^{K125A} and CarD^{K90A} are not expected to alter the concentration of CarD in *M. smegmatis* or *Mtb*, so this smaller effect is reflective of differences in regulation by CarD in the two organisms. In *Mtb*,

there was substantial correlation between the genes deregulated in CarD^{K125A} and CarD^{R47E}. This is not reflected in the *M. smegmatis* data set, where only 38 transcripts or 17% of the genes deregulated in the CarD^{K90A} strain are also deregulated in the CarD^{R25E} strain (Fig. 7D).

Genomic context correlates with CarD regulation in *M. smegmatis.* Transcription start sites have not been mapped in *M. smegmatis* so we were unable to identify promoters. We therefore looked for other possible determinants of CarD gene expression. Genomic localization can affect gene expression through differences in DNA packing and promoter accessibility. To identify a possible role for genomic context in CarD mediated regulation, we normalized reads in each of the CarD mutant strains to the CarD^{WT} strain and plotted these by genomic location to determine if there is a correlation between the genomic localization of a transcript and its regulation by CarD (Fig. 8A). We found genomic islands that are enriched for downregulated transcripts compared to the genome as a whole in both CarD^{R25E} and CarD^{K90A} and genomic islands enriched for upregulated transcripts compared to the genome as a whole in CarD^{K90A}. These findings suggest a potential role for genomic context in CarD regulation. Interestingly, while we there are very few significantly downregulated transcripts shared between both mutants, we find that the down regulated genomic islands are shared between the two mutants, suggesting that this effect of genomic context may affect both mutants.

To further investigate the possibility that genomic context can cause down regulation of transcripts in both CarDR25E and CarDK90A, we examined the 33 *M. smegmatis* transcripts downregulated in both the CarD^{R25E} and CarD^{K90A} strains (Fig. 7D). In line with a possible role for genomic context in deregulation of these shard transcripts, we found that 25 of these transcripts are encoded between *msmeg_3926* and *msmeg_3955*. Downregulation of genes in this region in both CarD^{R25E} and CarD^{K90A} was confirmed by quantitative real-time PCR (qRT-PCR)

(Fig. 8B). There are multiple operons encoded in this region and the predicted function of genes encoded here do not suggest a common function suggesting that this region may be downregulated due to genomic context rather than the transcripts encoded. In fact, every single gene in this region trends toward downregulation in both mutants but some transcripts did not meet the significance cut off in our RNA-seq analysis. Additionally, this region of the genome encodes 3 of the 5 most down downregulated transcripts in either mutant strain.

We have previously performed a screen for transposon mutants in *M. smegmatis* that restore CarD^{WT}-like expression of *lacZ* in the CarD^{R25E} mutant when expression was driven by rRNA operon promoters. Of the 551 hits found in this screen, 459 hits were mutants of *espR*, a nucleoid associated protein (unpublished data, (141)). This class of proteins bind DNA and can influence transcription via occlusion of promoters, bridging DNA, wrapping DNA, or bending DNA (142). They can also constrain supercoiling and compaction of DNA. To test is EspR was influencing regulation of *msmeg_3926-msmeg_3955*, we compared regulation of these genes in the CarD^{WT} and CarD^{R25E} strains in both a *espR* encoding and a *ΔespR* background by qRT-PCR (Fig 8C). Preliminary analysis suggests that deletion of *espR* may restore expression of these transcripts in the CarD^{R25E} background, although *ΔespR* also increases expression of these genes in the CarD^{WT} background. To determine if *ΔespR* increases expression of transcripts in the CarD^{R25E} background at all promoter, we performed qRT-PCR on a selection of genes deregulated in the CarD^{R25E} strain at other locations in the genome (Fig. 8D). At promoters from other regions of the genome *ΔespR* does not restore expression to CarD^{WT} levels. This supports a hypothesis in which is it a specific local effect of EspR on genes between *msmeg_3926-msmeg_3955* that is causing downregulation of these genes in the CarD mutants and indicating that CarD's activity can be affected by the presence of nucleoid associated proteins.

Genomic context correlates with CarD-mediated regulation in *Mtb*. Because genomic context may be a determinant of CarD regulation in *M. smegmatis*, we wanted to determine if we see a similar effect in *Mtb*. As in *M. smegmatis*, we normalized reads in each of the CarD mutant strains to the $CarD^{WT}$ strain and plotted these by genomic location (Fig. 9). This analysis clearly shows genomic islands where transcripts are more likely to be upregulated in the $CarD^{R47E}$, $CarD^{I27F}$, and $CarD^{I27W}$ strains suggesting these mutations may be influenced by the genomic context of the transcripts. Unlike in M. smegmatis, the genomic islands of deregulated genes in Mtb are not common shared between different classes of mutants. Interestingly, we see no evidence of an effect on genomic context in the mutant of the DNA binding domain suggesting that the effect of genomic context on CarD activity may be dependent on CarD's interaction with DNA and that this regulatory effect is lost when CarD's interaction with DNA is mutated.

Discussion

In this study, we examined CarD-mediated transcriptional regulation of throughout the genome in both *Mtb* and *M. smegmatis*. In *Mtb*, we found that modulating the concentration of CarD, affecting the affinity of CarD for RNAP, or mutating the DNA binding motif of CarD all have global effects on the transcriptome, together causing significant deregulation of 78% of the *Mtb* transcriptome. We found transcripts that we upregulated in loss of function CarD mutants, which suggests that $CarD^{WT}$ can repress or activate transcription in a transcript dependent manner. This is the first study to find evidence supporting $CarD^{WT}$-mediated repression of transcription. Other transcription factors that that regulate transcription through stabilization of the RNAP-promoter complexes have been shown to repress transcription by hyper-stabilizing these complexes, preventing promoter escape or transition to the elongation complex (124) so it is reasonable to suggest that CarD may repress transcription from some promoters. In *M.*

smegmatis, a much smaller portion of the genome is deregulated by decreasing CarD's affinity for RNAP or DNA suggesting there may be differences in the role of CarD-mediated regulation in the two organisms.

Our analysis revealed large effects on the transcriptome of the bacterium dependent of the concentration of CarD *in vivo*, with 52% of the transcriptome significantly deregulated when two strains of *Mtb* expressing different concentrations of $CarD^{WT}$. Additionally, changes in the cellular concentration of CarD resulting from an increased affinity of CarD for RNAP may be a major driver of transcript deregulation in these strains (Table 2). CarD was initially identified as gene upregulated in response to numerous different stress stimuli suggesting that mycobacteria may manipulate the concentration of CarD to globally regulate transcription in some environments.

We were able to correlate CarD-mediated regulation of a transcript to its promoter sequence by analyzing -10 promoter motifs of transcripts with known TSS in *Mtb*. The most common -10 promoter motif in *Mtb* is a TANNNT motif and, prior to this study, TANNNT motif promoters are the only type of promoter analyzed to determine the effect of CarD on regulation. By analyzing the correlations between deregulation in CarD mutants and promoter motifs we found that transcripts lacking a recognizable -10 promoter motif were upregulated by decreasing CarD's interactions with RNAP or DNA or by increasing CarD's cellular concentration over that in the $CarD^{WT}$ strain (Fig. 6AB).

Based on structural studies, it has been suggested that the interaction between W85 in CarD (W86 in the *Thermus* CarD) and a T at the -12 position of the promoter is important for transcriptional regulation by CarD (86, 87). While a strain singly expressing $CarD^{W85A}$ is not viable in *Mtb*, we have shown *in vitro* that mutating the interaction with RNAP or DNA has a

similar effect on regulation of transcription as the CarD^{W85A} mutant (108, 109). Through analysis of the promoters deregulated greater than 1.5-fold in CarD^{R47E} and/ or CarD^{K125A} we found no evidence for an increased rate of encoding a T at the -12 base than promoters of transcripts not deregulated at least 1.5-fold (40.67% and 40.86% of promoters have a T at the -12 position in deregulated and not deregulated respectively). The structural analysis was based on analysis of a promoter activated by CarD^{WT} containing a TANNNT motif in which the upstream T occurred at -12. This T in the TANNNT motif varies from -6 to -20 throughout the genome when the first 20 bp upstream of the promoter are analyzed for promoter motifs and is not more likely to occur at -12 in promoters containing a TANNNT motif that are deregulated by alterations in CarD than those that are not deregulated in our analysis (46.1% and 47.5%, respectively, of transcripts with a TANNNT promoter motif that are or are not deregulated greater than 1.5-fold in at least one strain have the upstream T of the TANNNT motif at -12). An interaction between W85 and the upstream T of the TANNNT motif may be important for CarD-mediated activation of TANNNT motif promoters but there is no evidence to support the importance of its position relative to the +1 site of the transcript.

CarD is essential in both *M. smegmatis* and *Mtb* and the protein is highly conserved through mycobacteria. Both the *M. smegmatis* and the *Mtb* CarD function similarly in biochemical assays including immunoprecipitation experiments, electrophoretic mobility shift assays (EMSAs), and transcription assays. However, throughout the years of studying CarD, we have identified several differences between CarD's function and regulation in the two organisms. Firstly, *Mtb* has more restrictive requirements for CarD than *M. smegmatis* as *Mtb* only tolerates mutations in CarD that have a minor effect on the interaction with RNAP or that occur within the DNA binding motif but retain the interaction with DNA in an EMSA assay while *M. smegmatis*

can survive mutations the dramatically decrease the affinity for RNAP and abolish the interaction with DNA, including one strain which likely affects the structure and decreases the stability of CarD (101, 108). *M. smegmatis* can additionally tolerate mutation of the conserved tryptophan to an alanine (W85A) in the *M. smegmatis* CarD protein but not the *Mtb* CarD protein. *Mtb* cannot tolerate the W85A mutation in either protein background. Furthermore, we have identified differences in post transcriptional regulation of CarD between the two organisms. In *Mtb*, the concentration of the protein in the cells is dependent on CarD's affinity for RNAP but this mechanism of controlling the protein concentration is not conserved in *M. smegmatis* (Fig7A, (136)). Based on the RNA-seq experiments detailed in this study, the role of CarD in regulating transcription is also much more limited in *M. smegmatis* than *Mtb*. Future studies examining the role of CarD in regulation of non-ribosomal promoters should be cautious in generalizing findings in one organism to make conclusions about other CarD-containing strains.

One aspect of regulation of CarD's transcription regulatory activity that may be conserved in both *Mtb* and *M. smegmatis* is a potential role for genomic context in regulation of transcription. In both *Mtb* and *M. smegmatis*, there are regions in the genome that have an increased chance of deregulated transcripts as compared to the genome as a whole. In *M. smegmatis* there is preliminary data to support a role for the nucleoid associated protein EspR in regulation of at least one of these genomic islands of deregulation. This data should be confirmed in *M. smegmatis* and expression of genes in a CarD mutant should be analyzed in a *ΔespR* background in *Mtb* to determine if the EspR affects regulation of genomic islands of deregulation in *Mtb*.

This manuscript is the first investigation of the role of CarD, which is localized to promoters throughout the genome, on regulation of non-ribosomal promoters and the first to

investigate in depth the role of the cellular concentration of CarD on gene expression *in vivo*. Through this analysis we have dramatically advanced our understanding of the determinants and extent of CarD mediated transcriptional regulation.

Acknowledgements

We are grateful to Leah Imlay for her thoughtful review of this chapter. C.L.S. is supported by Grant GM107544 from the National Institutes of Health. A.L.G. is supported by the NIGMS Cell and Molecular Biology Training Grant GM007067 and the Stephen I. Morse Graduate Fellowship.

Table Legends:

Table 1: Description of strains used in the *Mtb* RNA-seq analysis including the allele of CarD encoded, the effect of the mutation, and the concentration of CarD in the cell.

Table 2: Relative contributions of affinity and concentration to gene deregulation in the high affinity mutants. All possible expressions of genes in CarD ^{I27F} or CarD^{I27W} relative to the CarD^{WT} and WT *Mtb* strains, the number of genes in each category, and the regulatory scheme these expression patterns correspond to. Transcripts deregulated in the comparison of WT *Mtb* to CarD^{WT} are only included if also deregulated in the relevant high affinity mutant.

Table 3: Description of strains used in the *M. smegmatis* RNA-seq analysis including the allele of CarD encoded, the effect of the mutation, and the concentration of CarD in the cell.

Table 1: Description of the *Mtb* strains used in this study.

Strain	CarD Allele	Effect of Mutation	Cellular concentration of CarD
WT *Mtb*	$CarD^{WT}$	n/a	+++++
$CarD^{WT}$	$CarD^{WT}$	n/a	++
$CarD^{I27F}$	$CarD^{I27F}$	Higher affinity for RNAP	+++
$CarD^{I27W}$	$CarD^{I27W}$	Extrememly high affinity for RNAP	++++
$CarD^{R47E}$	$CarD^{R47E}$	Lower affinity for RNAP	+
$CarD^{K125A}$	$CarD^{K125A}$	Mutation in the DNA binding motif	++

Table 2: Relative contributions of affinity and concentration to gene deregulation in the high affinity mutants

			# of deregulated transcripts with this regulatory pattern		
Mutant to CarD^{WT}	**Mutant to WT *Mtb***	**WT Mtb to CarD^{WT}**	**CarD^{I27F}**	**CarD^{I27W}**	**Regulatory Scheme**
Down	No Change	No Change	12	32	
Down	No Change	Down	220	323	**C1:** Affinity and concentration contribute equally and have the same effect on regulation
Down	No Change	Up	0	0	
Down	Up	No Change	0	0	
Down	Up	Down	29	35	**B/ D1:** Concentration is the sole (B) or dominant (D1) driver of deregulation
Down	Up	Up	0	0	
No Change	Up	No Change	41	51	
No Change	Up	Down	201	180	**C2:** Affinity and concentration contribute equally and have opposeite effects on regulation
No Change	Up	Up	0	0	
Up	No Change	No Change	48	80	
Up	No Change	Down	0	0	
Up	No Change	Up	373	459	**C1:** Affinity and concentration contribute equally and have the same effect on regulation
Up	Down	No Change	1	1	
Up	Down	Down	0	0	
Up	Down	Up	14	17	**B/ D1:** Concentration is the sole (B) or dominant (D1) driver of deregulation
No Change	Down	No Change	48	57	
No Change	Down	Down	0	0	
No Change	Down	Up	196	212	**C2:** Affinity and concentration contribute equally and have opposeite effects on regulation
Down	Down	No Change	44	89	**A:** Affinity is the sole driver of deregulation
Down	Down	Down	17	42	**D2:** Affinity and concentration contribute, affinity is the dominant, and both have the same effect on regulation

			# of deregulated transcripts with this regulatory pattern		
Mutant to CarD^{WT}	**Mutant to WT *Mtb***	**WT Mtb to CarD^{WT}**	**CarD^{I27F}**	**CarD^{I27W}**	**Regulatory Scheme**
Down	Down	Up	10	14	**D3:** Affinity and concentration contribute, affinity is dominant, and they have opposite effects on regulation
Up	Up	No Change	91	136	**A:** Affinity is the sole driver of deregulation
Up	Up	Down	25	32	**D3:** Affinity and concentration contribute, affinity is dominant, and they have opposite effects on regulation
Up	Up	Up	28	73	**D2:** Affinity and concentration contribute, affinity is the dominant, and both have the same effect on regulation

Table 3: Description of the *M. smegmatis* strains used in this study.

Strain	CarD Allele	Effect of Mutation	Cellular concentration of CarD
$CarD^{WT}$	$CarD^{WT}$	n/a	++
$CarD^{R25E}$	$CarD^{R25E}$	Lower affinity for RNAP	++
$CarD^{K90A}$	$CarD^{K90A}$	Lower affinity for DNA	++

Figure Legends:

Figure 1: Transcripts deregulated in strains of *Mtb* expressing loss of function CarD mutants when compared to the $CarD^{WT}$ strain.

A-B. Pie charts showing regulation of transcripts upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ strain in the *ΔcarD attB::tet-carD* strain expressing $CarD^{R47E}$ (A) or $CarD^{K125A}$ (B). The number of transcripts in each category is shown.

C. Venn diagram comparing genes significantly deregulated in the *ΔcarD attB::tet-carD* strain expressing $CarD^{R47E}$ (pink) or $CarD^{K125A}$ (green) upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$.

Figure 2: Transcripts deregulated in strains of *Mtb* expressing mutants of CarD with higher affinity for RNAP when compared to the $CarD^{WT}$ strain.

A-B. Pie charts showing regulation of transcripts upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ strain in the *ΔcarD attB::tet-carD* strain expressing $CarD^{I27F}$ (A) or $CarD^{I27W}$ (B). The number of transcripts in each category is shown.

C. Venn diagram comparing genes significantly deregulated in the *ΔcarD attB::tet-carD* strain expressing $CarD^{I27F}$ (purple) or $CarD^{I27W}$ (blue) upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$.

Figure 3: Increasing and decreasing CarD's affinity for RNAP have different effects on the transcriptome in *Mtb*.

A-B. Venn diagram comparing genes significantly deregulated in the *ΔcarD attB::tet-carD* strain expressing $CarD^{R47E}$ (pink) to genes deregulated in the *ΔcarD attB::tet-carD* strain expressing $CarD^{I27F}$ (A, purple) or $CarD^{I27W}$ (B, purple) upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$.

Figure 4: Increasing the concentration of CarD^{WT} has phenotypic consequences in *Mtb*.

A. The ratio of *carD* transcript levels in exponential cultures of the *ΔcarD attB::tet-carD* strain expressing CarD^{WT} or WT *Mtb* or the to levels in CarD^{WT} when grown in Sauton's media. Transcript levels were determined using qRT-PCR and normalized to *sigA*.

B. Western blot analysis of two biological replicates of lysates from exponentially growing *Mtb ΔcarD attB::tet-carD* strain expressing CarD^{WT} or WT *Mtb* in Sauton's media. Membranes were blotted with a monoclonal antibody against RNAP-β (top) or a polyclonal antibody against CarD (bottom).

C. Representative growth curves of the *ΔcarD attB::tet-carD* strain expressing CarD^{WT} (circles) or WT *Mtb* WT *Mtb* (squares) in Sauton's media.

D. Doubling times of the *ΔcarD attB::tet-carD* strain expressing CarD^{WT} or WT *Mtb* in Sauton's media.

E. Ratio of 16S rRNA levels in exponential cultures of the *ΔcarD attB::tet-carD* strain expressing CarD^{WT} or WT *Mtb* to levels in CarD^{WT} when grown in Sauton's media. Transcript levels were determined using qRT-PCR and normalized to *sigA*.

F. Dot plot of colony forming units (CFU) in the lungs of C57BL/6J mice infected with the *ΔcarD attB::tet-carD* strain expressing CarD^{WT} (circles) or WT *Mtb* WT *Mtb* (squares). Each time point is the mean ± SEM of data from at least 6 mice per strain.

Each graph shows the mean ± SEM of data from at least three replicates. Significance was determined by t-test. *, $p \leq 0.05$; **, $p \leq 0.01$.

Figure 5: Increasing the concentration of CarD^{WT} affects regulation of half of the transcriptome in *Mtb*.

A-B. Pie charts showing regulation of transcripts in WT *Mtb* upon comparison to the $CarD^{WT}$ strain. The number of transcripts in each category is shown.

Figure 6: Transcriptional regulation by CarD is affected by the -10 promoter motif.

A. The percentage of transcripts with a mapped TSS that have an upstream TANNNT (dark blue), GNNANNNT (light blue), or no detectable (red) promoter motif in the genome as a whole, in transcripts not deregulated by alterations in CarD, or significantly deregulated by alterations in CarD with increasingly stringent fold-change cut offs from left to right.

B. The percentage of transcripts with a mapped TSS that have an upstream TANNNT (dark blue), GNNANNNT (light blue), or no detectable (red) promoter motif in the genome as a whole, transcripts not deregulated 1.5-fold in any tested strain, deregulated 1.5-fold in any tested strain, transcripts deregulated at least 1.5-fold by at least one tested alteration in CarD, or up or down regulated greater than 1.5 fold in WT *Mtb*, or *ΔcarD attB::tet-carD* strain expressing $CarD^{I27F}$, $CarD^{I27W}$, $CarD^{R47E}$, or $CarD^{K125A}$. Statistical significance was determined by a Chi-square test.

D. Graph showing the number of different strains a transcript is significantly deregulated in by -10 promoter motifs. Statistical significance was analyzed by ANOVA and Tukey's multiple comparison test.

*, $p \leq 0.05$; **, $p \leq 0.01$; ***, or $p \leq 0.001$. All comparisons were tested and only significantly different comparisons are shown.

Figure 7: Loss of function mutations in CarD effect a smaller percentage of the genome in *M. smegmatis* than *Mtb*.

A. Western blot analysis of two biological replicates of lysates from exponentially growing *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ or $CarD^{R25E}$. Membranes were blotted with a monoclonal antibody against RNAP-β (top) or a polyclonal antibody against CarD (bottom).

B-C. Pie charts showing regulation of transcripts upon comparison to the *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ strain in the *ΔcarD attB::tet-carD* strain expressing $CarD^{R25E}$ (A) or $CarD^{K90A}$ (B). The number of transcripts in each category is shown.

D. Venn diagram comparing genes significantly deregulated in *ΔcarD attB::tet-carD* strain expressing $CarD^{R25E}$ (pink) or $CarD^{K90A}$ (green) upon comparison to the *ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$.

Figure 8: CarD's regulatory activity may be affected by genomic context and the nucleoid associated protein EspR in *M. smegmatis*.

A. Reads in *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{R25E}$ (red/gold) or $CarD^{K90A}$ (blue/aqua) were normalized to read counts in the *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$ strain and plotted based on their position in the genome (nucleotide positions are numbered). Transcripts with lower read counts in the mutant are graphed in gold or aqua and transcripts with higher read counts in the mutant are graphed in red or blue. Genomic islands of transcripts downregulated in both mutants are designated by an *. Genomic islands of transcripts specifically upregulated in $CarD^{K90A}$ are designated by a #.

B. Ratio of select transcripts between *msmeg_3926* and *msmeg_3955* in exponential cultures of the *M. smegmatis ΔcarD attB::tet-carD* strain expressing $CarD^{WT}$,

CarD^{R25E} or CarD^{K90A} to levels in CarD^{WT}. Transcript levels were determined using qRT-PCR and normalized to *sigA*.

C. Ratio of select transcripts between *msmeg_3926* and *msmeg_3955* in exponential cultures of the *M. smegmatis ΔcarD attB::tet-carD* strain expressing CarD^{WT} or CarD^{R25E} in the presence or absence of *espR* to levels in CarD^{WT} with *espR* . Transcript levels were determined using qRT-PCR and normalized to *sigA*.

D. Ratio of select transcripts not encoded between *msmeg_3926* and *msmeg_3955* in exponential cultures of the *M. smegmatis ΔcarD attB::tet-carD* strain expressing CarD^{WT} or CarD^{R25E} in the presence or absence of *espR* to levels in CarD^{WT} with *espR* . Transcript levels were determined using qRT-PCR and normalized to *sigA*.

Figure 9: CarD's regulatory activity may be affected by genomic context in *Mtb*.

A. Reads in *Mtb ΔcarD attB::tet-carD* strain expressing CarD^{R47E} (periwinkle/red), CarD^{K125A} (magenta/ yellow), CarD^{I27W} (blue/green), or CarD^{I27F} (navy/ green) were normalized to read counts in the *Mtb ΔcarD attB::tet-carD* strain expressing CarD^{WT} strain and plotted based on their position in the genome (nucleotide positions are numbered). Transcripts with lower read counts in the mutant are graphed in periwinkle, magenta, blue or navy and transcripts with higher read counts in the mutant are graphed in red, yellow, or green. Genomic islands of transcripts downregulated in both high affinity mutants are designated by an *. Genomic islands of transcripts specifically upregulated in CarD^{R47E} are designated by a #.

Figure 1

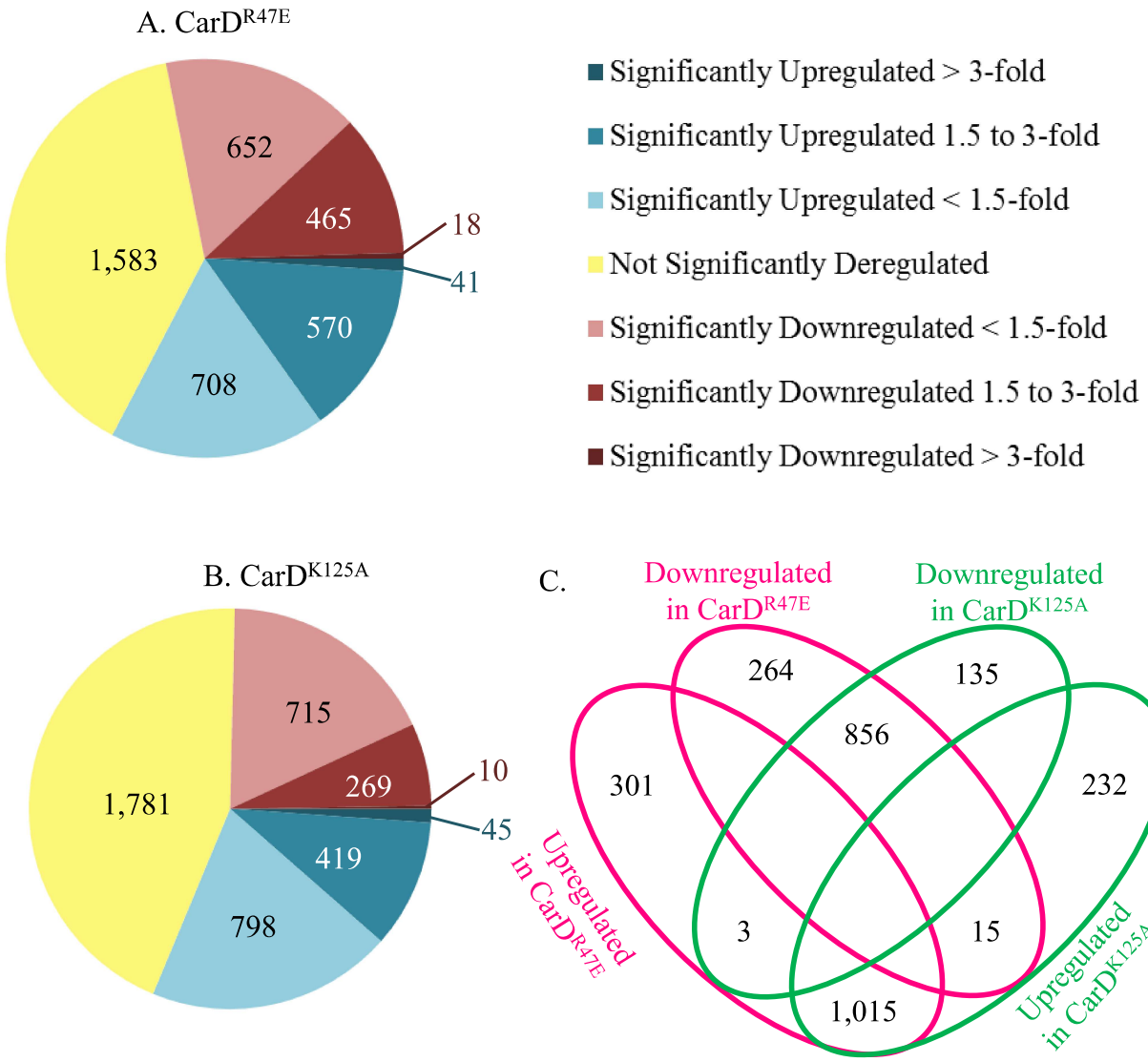
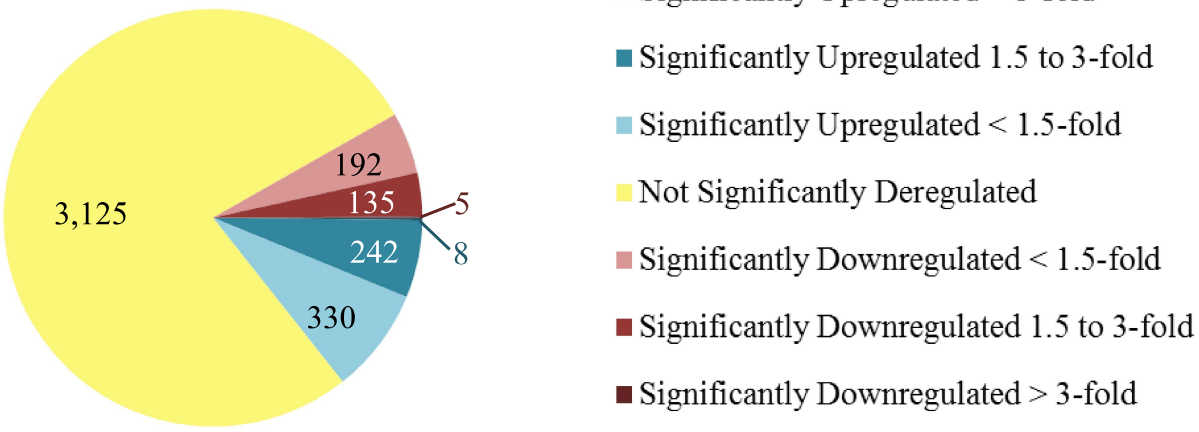
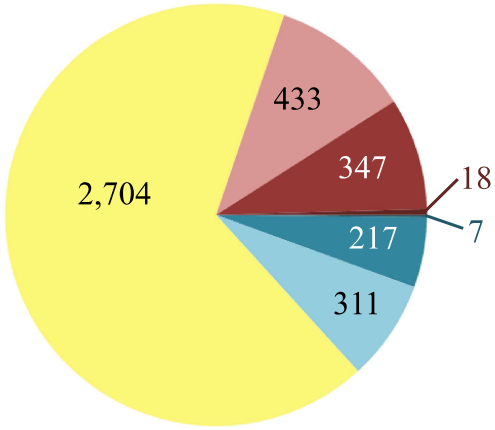

A. CarD^R47E
652
465
18
1,583
41
570
708
Significantly Upregulated > 3-fold
Significantly Upregulated 1.5 to 3-fold
Significantly Upregulated < 1.5-fold
Not Significantly Deregulated
Significantly Downregulated < 1.5-fold
Significantly Downregulated 1.5 to 3-fold
Significantly Downregulated > 3-fold
B. CarD^K125A
715
269
10
1,781
45
419
798
C.
Downregulated in CarD^R47E
Downregulated in CarD^K125A
264
135
856
301
232
Upregulated in CarD^R47E
3
15
1,015
Upregulated in CarD^K125A

Figure 2

A. CarD^{I27F}



B. CarD^{I27W}



C.

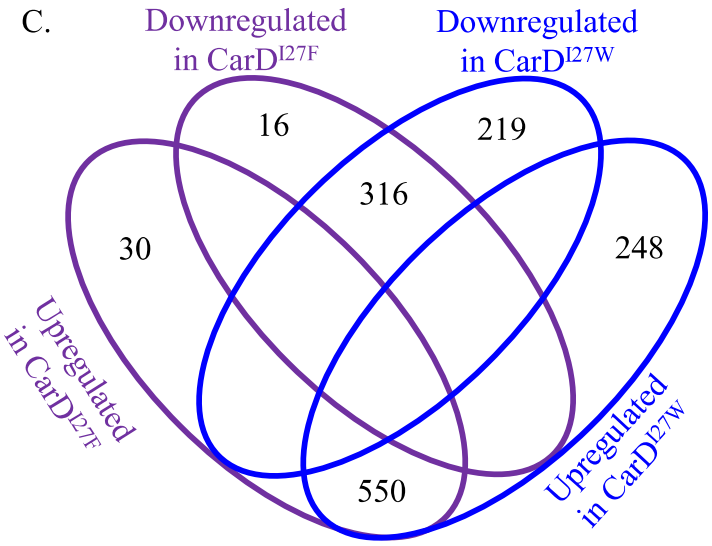
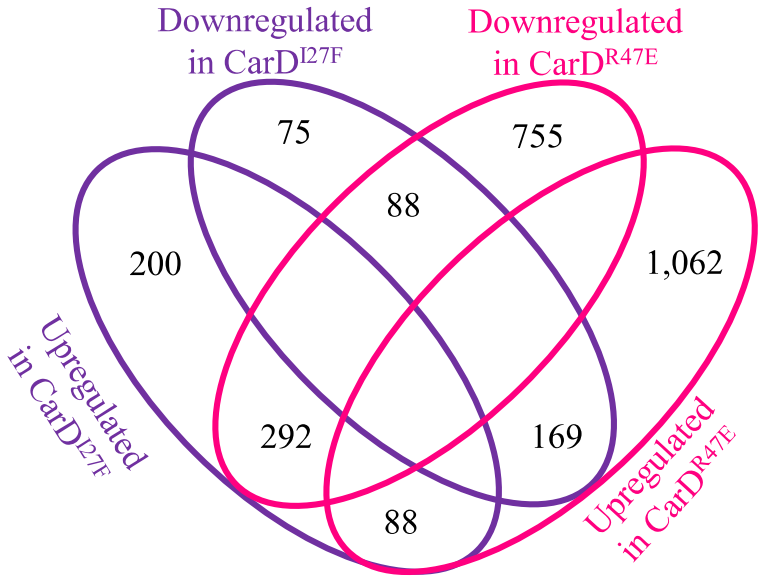


Figure 3

A.


Downregulated in CarD^I27F
Downregulated in CarD^R47E
75
755
88
200
1,062
Upregulated in CarD^I27F
292
169
88
Upregulated in CarD^R47E

B.

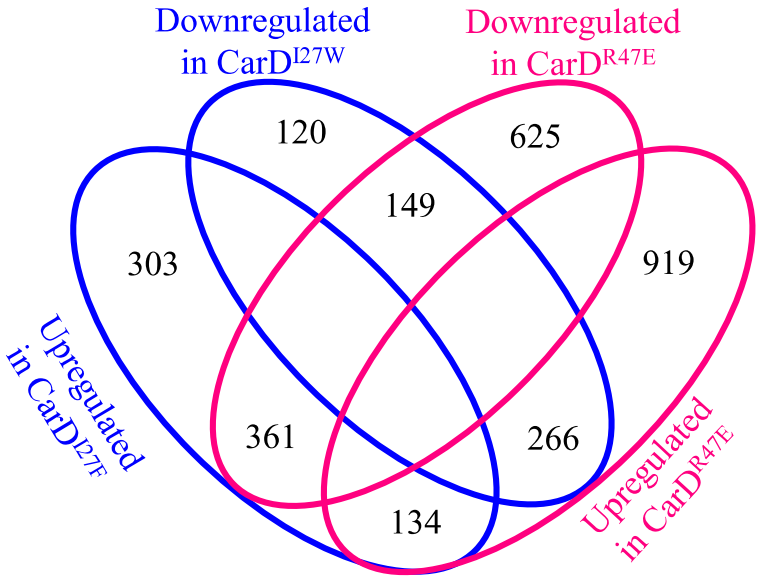

Downregulated in CarD^I27W
Downregulated in CarD^R47E
120
625
149
303
919
Upregulated in CarD^I27F
361
266
134
Upregulated in CarD^R47E

Figure 4

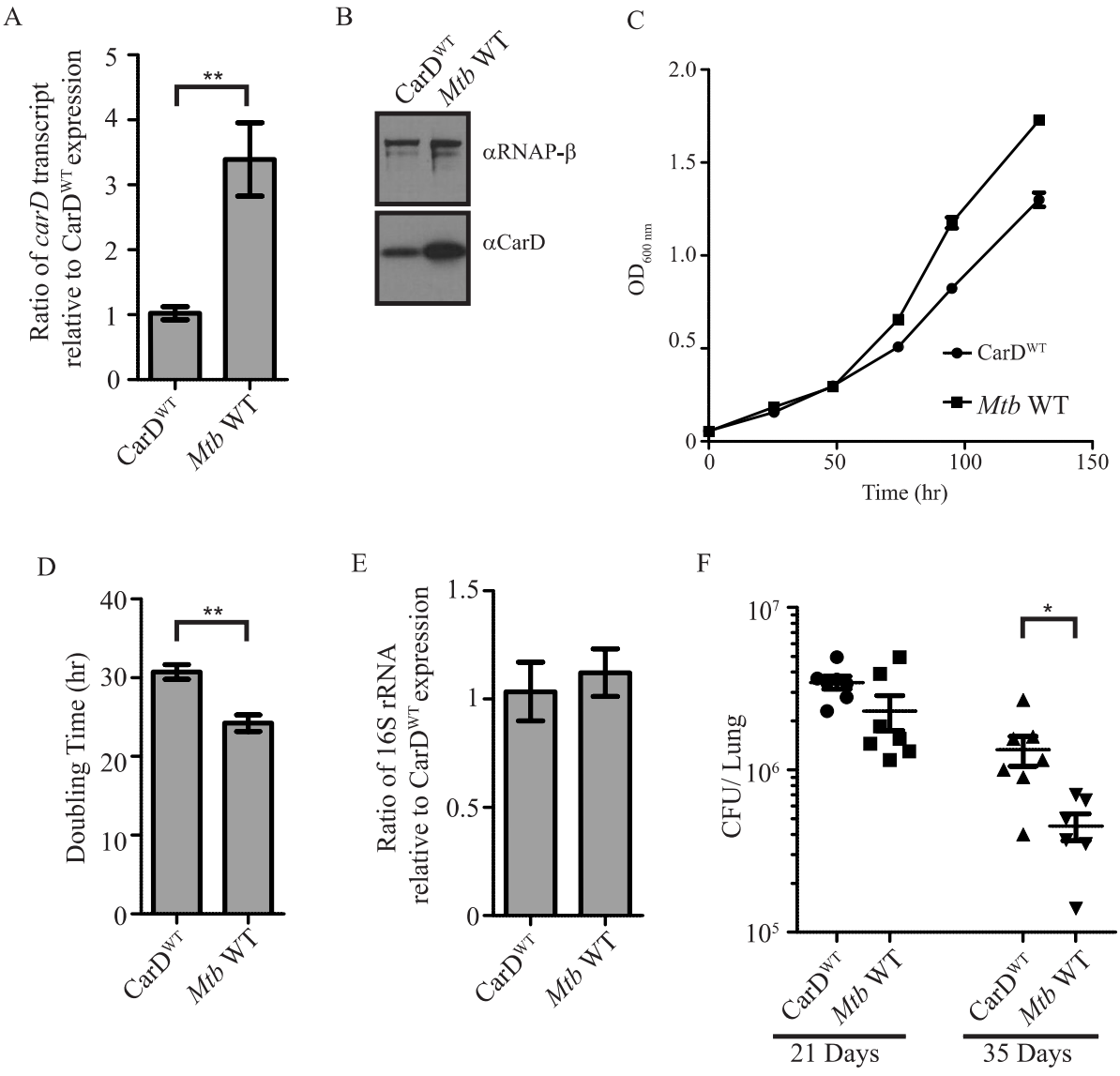

A
Ratio of carD transcript relative to CarD^WT expression
**
CarD^WT
Mtb WT
B
CarD^WT
Mtb WT
αRNAP-β
αCarD
C
OD600 nm
CarD^WT
Mtb WT
Time (hr)
D
Doubling Time (hr)
**
CarD^WT
Mtb WT
E
Ratio of 16S rRNA relative to CarD^WT expression
CarD^WT
Mtb WT
F
CFU/ Lung
*
CarD^WT
Mtb WT
21 Days
CarD^WT
Mtb WT
35 Days

Figure 5

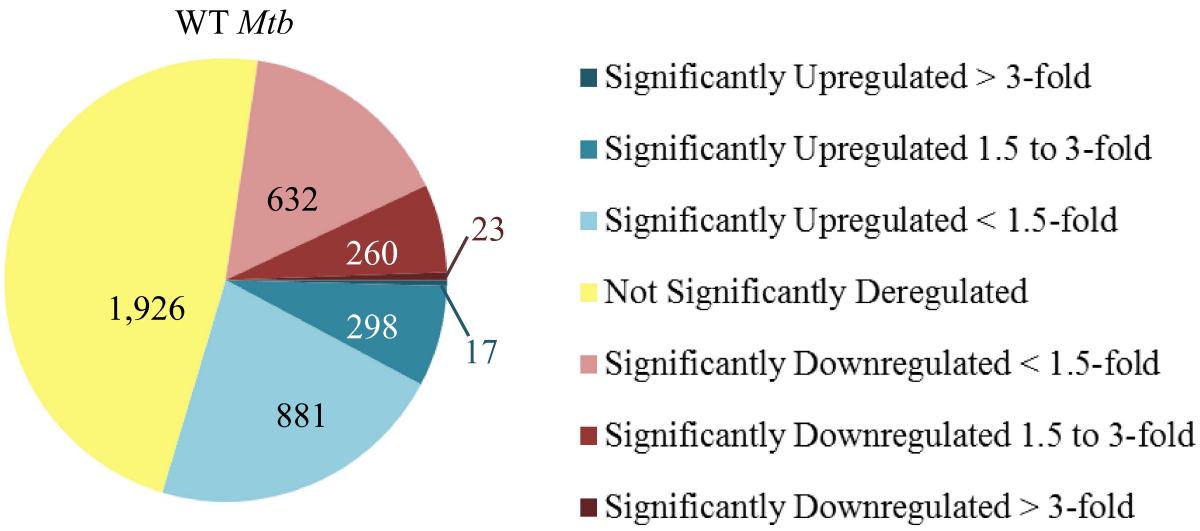

WT *Mtb*
632
260
23
1,926
298
17
881
Significantly Upregulated > 3-fold
Significantly Upregulated 1.5 to 3-fold
Significantly Upregulated < 1.5-fold
Not Significantly Deregulated
Significantly Downregulated < 1.5-fold
Significantly Downregulated 1.5 to 3-fold
Significantly Downregulated > 3-fold

Figure 6

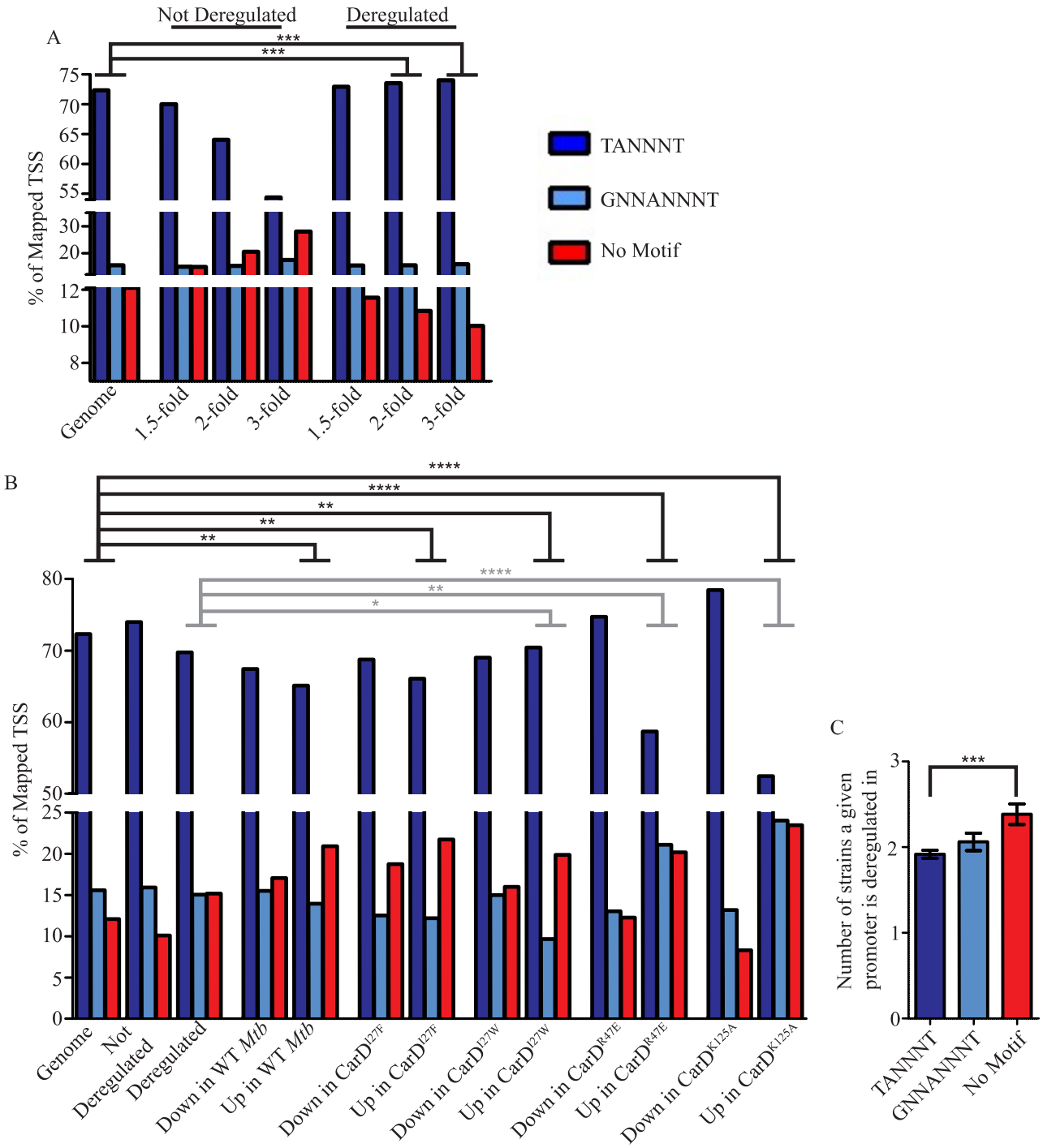
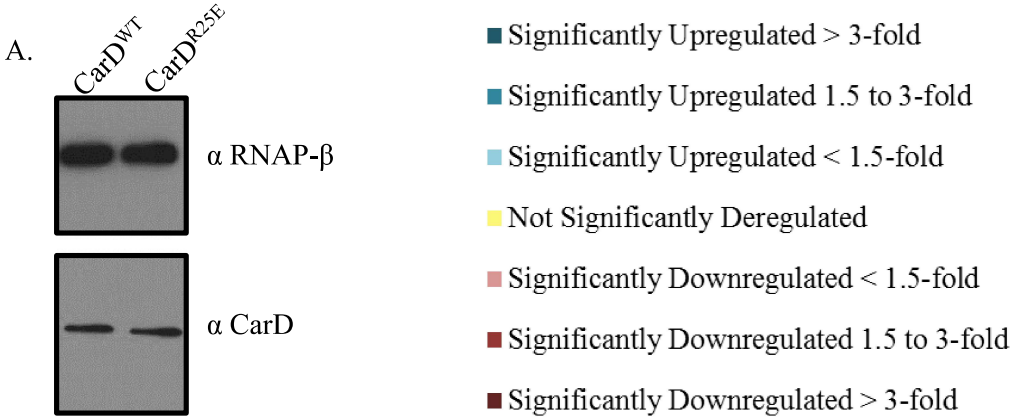
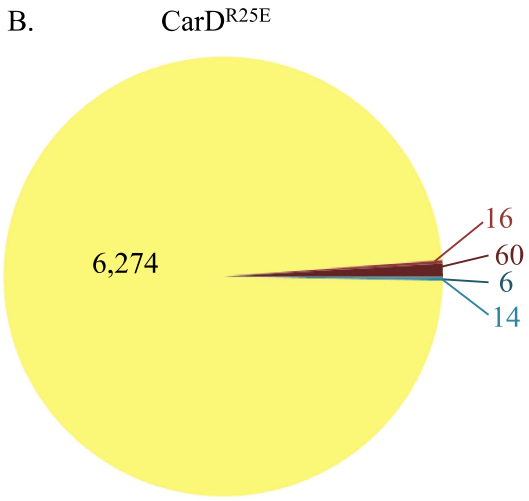
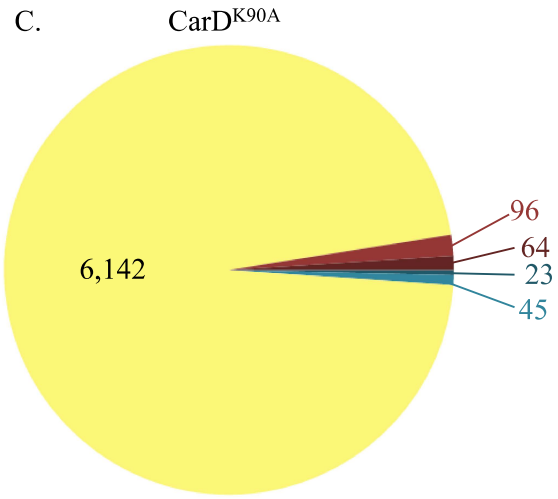

A
Not Deregulated
Deregulated
***
***
% of Mapped TSS
75
70
65
60
55
30
20
12
10
8
Genome
1.5-fold
2-fold
3-fold
1.5-fold
2-fold
3-fold
TANNNT
GNNANNNT
No Motif
B
****
****
**
**
**
*
% of Mapped TSS
80
70
60
50
25
20
15
10
5
0
Genome
Not Deregulated
Deregulated
Down in WT Mtb
Up in WT Mtb
Down in CarD I27F
Up in CarD I27F
Down in CarD I27W
Up in CarD I27W
Down in CarD R47E
Up in CarD R47E
Down in CarD K125A
Up in CarD K125A
C
***
Number of strains a given promoter is deregulated in
3
2
1
0
TANNNT
GNNANNNT
No Motif

Figure 7


A.
CarD^WT
CarD^R25E
α RNAP-β
α CarD
Significantly Upregulated > 3-fold
Significantly Upregulated 1.5 to 3-fold
Significantly Upregulated < 1.5-fold
Not Significantly Deregulated
Significantly Downregulated < 1.5-fold
Significantly Downregulated 1.5 to 3-fold
Significantly Downregulated > 3-fold


B.
CarD^R25E
6,274
16
60
6
14


C.
CarD^K90A
6,142
96
64
23
45

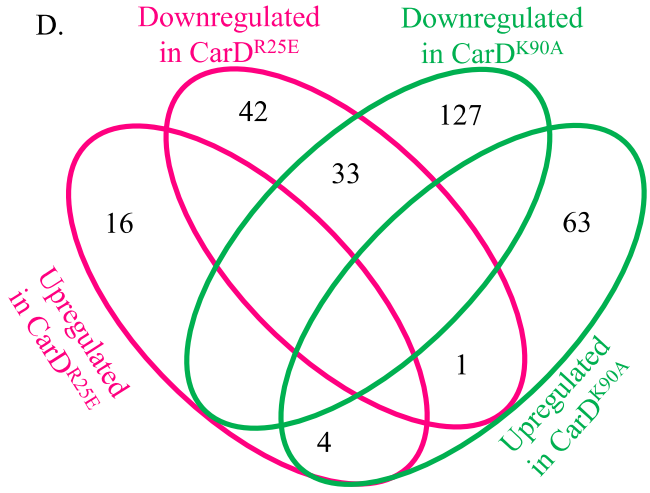

D.
Downregulated in CarD^R25E
Downregulated in CarD^K90A
Upregulated in CarD^R25E
Upregulated in CarD^K90A
42
127
33
16
63
1
4

Figure 8

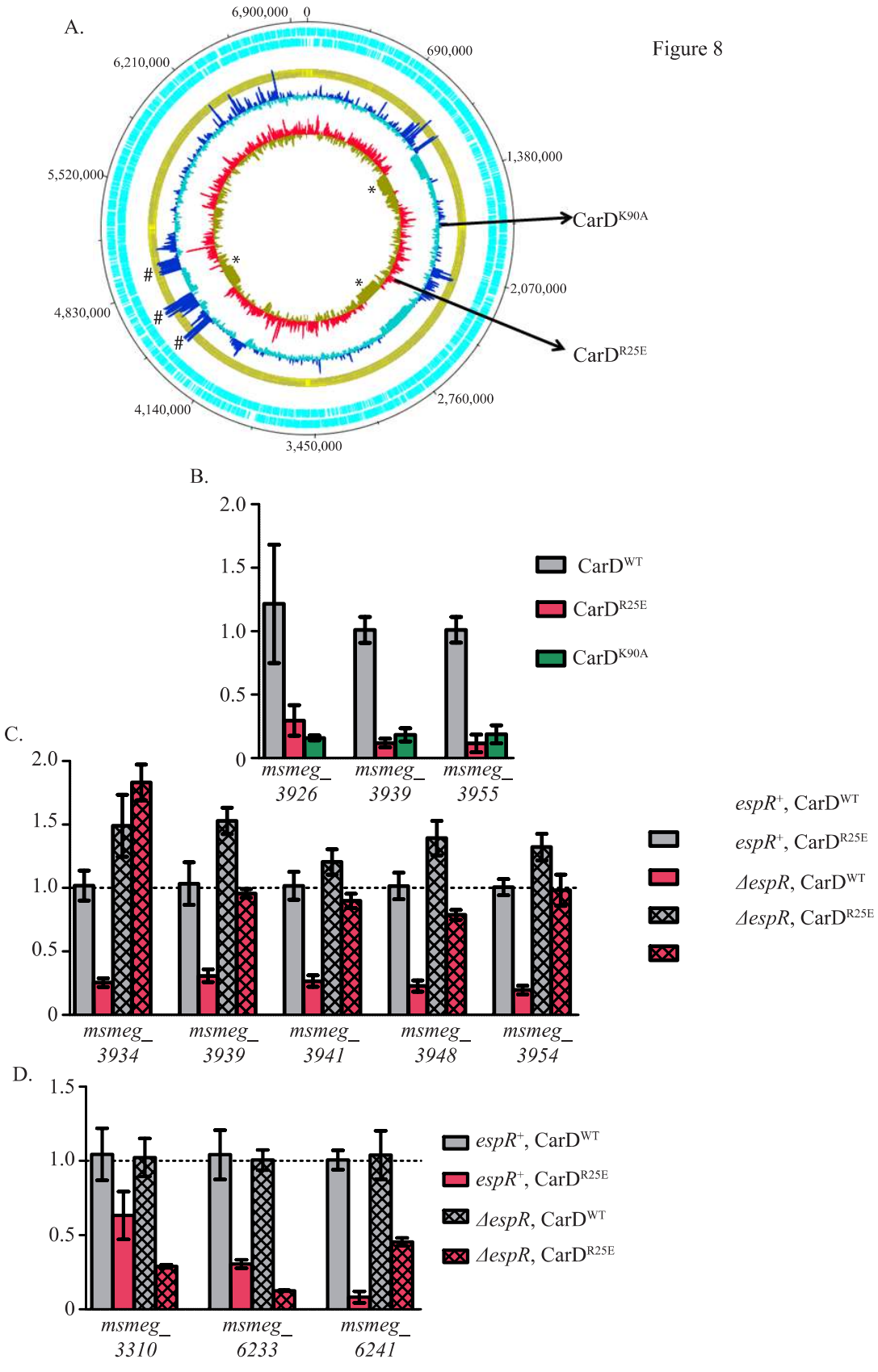
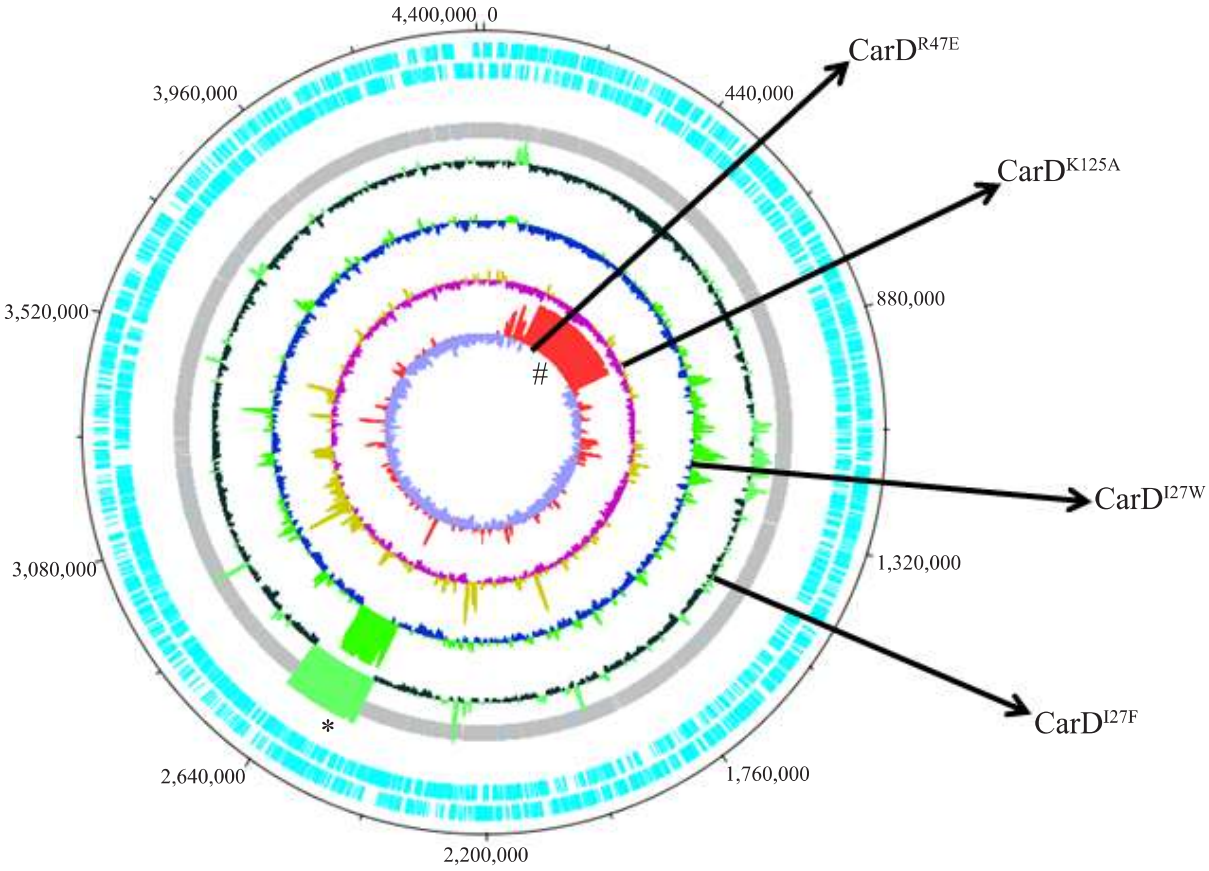

A.
6,900,000
0
690,000
1,380,000
2,070,000
2,760,000
3,450,000
4,140,000
4,830,000
5,520,000
6,210,000
CarD^K90A
CarD^R25E
B.
CarD^WT
CarD^R25E
CarD^K90A
msmeg_ 3926
msmeg_ 3939
msmeg_ 3955
C.
espR+, CarD^WT
espR+, CarD^R25E
ΔespR, CarD^WT
ΔespR, CarD^R25E
msmeg_ 3934
msmeg_ 3939
msmeg_ 3941
msmeg_ 3948
msmeg_ 3954
D.
espR+, CarD^WT
espR+, CarD^R25E
ΔespR, CarD^WT
ΔespR, CarD^R25E
msmeg_ 3310
msmeg_ 6233
msmeg_ 6241

Figure 9


4,400,000 0
440,000
880,000
1,320,000
1,760,000
2,200,000
2,640,000
3,080,000
3,520,000
3,960,000
CarD^R47E
CarD^K125A
CarD^I27W
CarD^I27F
#
*

Chapter 5: Conclusions

Ashley Garner

Major Findings:

Mycobacterium tuberculosis (*Mtb*) infects about a third of the world's population and causes more than a million death per year (4). To gain control of this epidemic we must develop new therapeutics, which requires a better understanding of unique mycobacterial physiology. It has recently become clear that the mycobacterial RNA polymerase (RNAP) differs substantially from the well-studied RNAP of *Escherichia coli*, upon which the majority of studies on transcription have been performed. Our group and others have shown that, in comparison to *E. coli*, the mycobacterial RNAP is deficient in its ability to form stable open complex (RP_o), an intermediate complex formed during transcription initiation (109, 123). Mycobacteria overcome this deficiency of their RNAP with essential accessory transcription regulators, including CarD, which represent promising targets for future therapeutics.

CarD was initially discovered as a gene upregulated in response to numerous physiologically relevant stresses (85). It is an essential, highly-conserved gene in mycobacteria with homologs in numerous, but not all, bacterial species (85–87). Importantly, there are no CarD homologs in eukaryotes. Initial studies on CarD focused on its N-terminal domain, which is homologous to the RNAP-interacting domain of the transcription repair coupling factor (TRCF, encoded by *mfd*). Through this domain, CarD interacts with the β1 domain of the β subunit of RNAP. CarD's interactions with RNAP suggested that CarD was likely to be a transcriptional regulator (Chapter 1 Fig. 1, (85, 101). To determine which stage of transcription may be regulated by CarD, chromatin immunoprecipitation sequencing (ChIP-seq) reactions were performed that determined that CarD is co-localized with the RNAP holoenzyme at promoters throughout the genome (Chapter 1 Fig. 2, (87)). This result suggests that CarD likely regulates transcription initiation. CarD's crystal structure was therefore modeled onto initiation

complexes of RNAP. In the resulting model, CarD's C-terminal domain, which is largely α-helical in structure and not homologous to any previously solved protein fold, was poised to interact with DNA just upstream of the -10 motif of the promoter (87). Through electrophoretic mobility shift assays (EMSAs) we demonstrated that this domain of CarD interacts with DNA via a conserved basic patch in a sequence-independent manner (Chapter 1 Fig. 4, (87)).

CarD's interactions with RNAP and DNA, as well as the activity of a highly conserved tryptophan, are required for optimal growth, antibiotic resistance, and pathogenesis. To determine the functional importance of individual activities of CarD to its physiological function, we developed and characterized a series of single point mutants in CarD that each target an individual activity of CarD. We identified two arginine residues (R25 and R47) in CarD that are important for CarD's interactions with RNAP. Mutation of either residue to glutamate decreases CarD's affinity for RNAP (Chapter 1 Fig. 1, (101)). Within the conserved basic patch, we identified two lysine residues (K90 and K125) that are both important for CarD's interaction with DNA. An alanine mutation at K90 ($CarD^{K90A}$) or a glutamate mutation at K125 ($CarD^{K125E}$) abolish CarD's ability to bind and shift DNA in an EMSA assay (Chapter 2 Fig. 1, (108)). An alanine mutant at K125 ($CarD^{K125A}$) retains the interaction with DNA in an EMSA assay, but has functional consequences for the physiology of the bacterium, suggesting this mutation may affect CarD's interaction with DNA *in vivo* (Chapter 2, (108)). Within the basic patch of CarD, we have additionally identified a highly conserved tryptophan residue (W85) that is not essential for DNA binding and is predicted to have an alternative activity in regulation of transcription. Specifically, structural studies predict this residue functions by wedging into the distorted minor groove at the upstream edge of the transcription bubble (86, 87). We constructed strains of *Mycobacterium smegmatis* that singly express $CarD^{WT}$, a CarD mutant with decreased

affinity for RNAP ($CarD^{R25E}$ or $CarD^{R47E}$), a CarD mutant of the DNA binding domain ($CarD^{K90A}$, $CarD^{K125A}$, or $CarD^{K125E}$), or a CarD mutant of the conserved tryptophan residue ($CarD^{W85A}$). In *Mtb*, only the CarD mutant with mild effects on the interaction with RNAP ($CarD^{R47E}$) and the mutant that retained the interaction with DNA *in vitro* ($CarD^{K125A}$) were viable, indicating that *Mtb* has more stringent requirements for CarD than *M. smegmatis* (Chapter 2). In both organisms, these mutation of CarD decreased the growth rate, indicating that CarD's interactions with RNAP and DNA as well as the activity of the conserved tryptophan residue are essential for CarD to regulate the transcription of genes required for wild-type (WT)-like growth (Chapter 1 Fig. 1, Chapter 2 Fig. 2, (101, 108)).

Because CarD is upregulated in response to numerous stresses (85), we then tested the effects of decreased affinities for RNAP or DNA or mutation of the conserved tryptophan on antibiotic sensitivity in *M. smegmatis*. During growth on agar plates, strains of *M. smegmatis* with any of these changes showed increased sensitivity to ciprofloxacin, rifampicin, and streptomycin, which target DNA replication, RNA synthesis, and protein synthesis, respectively (Chapter 1 Fig. 1, Chapter 2 Fig. 5, (101, 108)). This phenotype of general stress sensitivity indicates that mutations in CarD likely inhibit an upstream step required in response to numerous diverse stresses.

In addition to these stress, CarD is also transcriptionally upregulated in response to oxidative stress (85). In *M. smegmatis*, we found that CarD's interaction with RNAP and the activity of the conserved tryptophan were required for resistance to oxidative (H_2O_2) stress; however, in contrast to results seen during antibiotic stress experiments, mutations that decreased CarD's interaction with DNA did not sensitize bacteria to oxidative stress (Chapter 1 Fig. 1, Chapter 2 Fig. 3, (101, 108)). Importantly, the conserved tryptophan residue, which is required

for resistance to oxidative stress, is physically located within the DNA binding motif of CarD (Chapter 1 Fig. 3). The necessity of the tryptophan residue but not the surrounding basic patch for resistance to oxidative stress strongly supports the hypothesis that this residue has a physiological activity distinct from CarD's interaction with DNA. Mutation of the tryptophan residue is not viable in *Mtb* and therefore could not be tested; however, the *Mtb* CarD^{R47E} strain is much more sensitive to oxidative stress than the CarD^{WT} strain, again underlining the importance of CarD's interaction with RNAP (Chapter 1 Fig. 1, (101)). The variable requirement for individual CarD activities in response to oxidative stress indicates that some transcripts regulated by CarD only require a subset of CarD's activities and suggests that CarD's activity may be promoter-specific.

During mouse infections, *Mtb* strains expressing CarD^{K125A} or CarD^{R47E} reach similar bacterial burden in the lungs of the mice during acute infection, but are maintained at lower bacterial burden during persistent infection, in comparison to the CarD^{WT} strain. In the spleen, the bacterial burden of these strains is lower than the CarD^{WT} strain during both acute and persistent infection. These data suggest that CarD's interactions with RNAP and DNA are required for resistance to stress or virulence during infection (Chapter 1 Fig. 1, Chapter 2 Fig. 4, (101, 108)).

CarD requires all three of its functional domains to activate transcription from ribosomal RNA operons by stabilizing RP_o. During initial characterization of CarD, it was found that depletion of the CarD protein broadly deregulated transcripts associated with the translational machinery, including ribosomal RNA (rRNA) (85). Regulation of rRNA is an essential and tightly controlled process in all bacteria as rates of ribosomal protein transcription, ribosome biogenesis, and cell growth correlate to the levels of rRNA production (118–120). We

therefore focused our initial studies of transcriptional regulation by CarD on regulation of rRNA operons. In *Mtb* there is only one rRNA operon, *rrnA*, transcription of which is regulated by two promoters, P1 and P3 (116, 117). The homologous *rrnA* operon in *M. smegmatis* has P1 and P3 promoters and an additional P2 promoter (116, 117). In *lacZ* reporter experiments, we found that $CarD^{WT}$ activates transcription from *rrnA* promoters by a mechanism that is dependent on its interactions with RNAP and DNA as well as the activity of the conserved tryptophan (Chapter 2 Fig. 6, (108)).

To study the mechanism by which CarD activates transcription, we interrogated CarD's effect on transcription in a single round *in vitro* transcription assay that measures the stability of RNAP-promoter complexes. We found that the addition of $CarD^{WT}$ dramatically stabilizes RNAP-promoter complexes at rRNA promoters and this stabilizing effect is dramatically attenuated by mutations that decrease CarD's interaction with RNAP or DNA or mutation of the conserved tryptophan residue (Chapter 2 Fig. 7, (87, 108). Subsequently, researchers used $KMnO_4$ foot printing assays to show that the RP_o, specifically, is stabilized by addition of CarD (123).

Our understanding of the mechanism by which CarD stabilizes RP_o at rRNA promoters was advanced by a kinetic analysis of CarD transcriptional regulation at a rRNA promoter, performed by our collaborators in the Galburt lab (109). This study demonstrated that CarD interacts with RP_o with high affinity and acts to stabilizes the complex by reducing the reverse rate of collapse back to RP_c. CarD also increases the forward rate of promoter melting through a lower affinity interaction with RP_c. Importantly, I determined that the cellular concentrations of RNAP β, RNAP σ, and CarD in *Mtb* and are similar to concentrations used in the fluorescence

assay, suggesting that CarD both inhibits promoter collapse and enhances opening *in vivo* (Chapter2, Fig. 8B, (109)).

Increasing CarD's affinity for RNAP increases growth rate without affecting rRNA levels and decreases virulence in *Mtb*. To further dissect the relationship between CarD's interaction with RNAP, its regulation of rRNA, and the mycobacterial growth rate, we designed two CarD mutants with increased affinity for RNAP ($CarD^{I27F}$ and $CarD^{I27W}$). Both mutations have increased affinity for RNAP in immunoprecipitation and *biochemical* assays, though $CarD^{I27W}$ has a higher affinity for RNAP than $CarD^{I27F}$ does (Chapter 3 Fig. 1, (136)). When these mutants are expressed in *Mtb* they grow faster than the $CarD^{WT}$ strain (Chapter 3 Fig. 3, *carD* is expressed at similar levels transcriptionally in the $CarD^{WT}$, $CarD^{I27F}$, and $CarD^{I27W}$ strains, but protein concentrations increase with CarD's affinity for RNAP, with $CarD^{WT}$ having lower CarD concentrations than $CarD^{I27F}$, which in turn has less than $CarD^{I27W}$ (Chapter 3 Fig. 3, (136)). These data suggest that CarD, a known target of the Clp protease in mycobacteria (133), is regulated post-transcriptionally and may be protected from proteolytic degradation when associated with RNAP. Because mutations in CarD that decrease growth rate also decrease 16S rRNA content, we examined the levels of rRNA in these strains by quantitative real time PCR (qRT-PCR) (Chapter 2 Fig. 6, (87, 108). Surprisingly, the $CarD^{WT}$ strain and both strains expressing mutants of CarD with higher affinity for RNAP had similar levels of rRNA (Chapter 3 Fig. 3, (136)). These data indicate that the growth defect in the $CarD^{WT}$ strain was neither caused nor ameliorated by alterations in the rRNA content of the bacteria, thus decoupling rRNA content and growth rate in the CarD mutants. These studies reveal a new mechanism of regulating CarD activity through turnover of free protein and improve our understanding of the relationship between rRNA transcription, growth rate, and CarD activity.

To test if the restored growth rate of the CarD mutants with higher affinity for RNAP corresponded to increased fitness during infection, we infected C57BL/6J mice with WT *Mtb*, $CarD^{WT}$, $CarD^{I27F}$ or $CarD^{I27W}$. The CarD mutants with increased affinity for RNAP have decreased bacterial burden at 35 days post infection as compared to the $CarD^{WT}$ strain. Additionally, the $CarD^{I27W}$ strain, which has the highest affinity for RNAP, also has a lower bacterial burden at 21 days post infection, indicating that increasing CarD's affinity for RNAP is detrimental to virulence in *Mtb* (Chapter 3 Fig. 4, (136)).

Increasing CarD's affinity for RNAP increases the lifetime of RP_o. To investigate whether increasing the affinity of CarD for RNAP affects the stability of transcription-competent RNAP-promoter complexes, we performed a multi-round *in vitro* aborted transcription assay (123). Increasing CarD's affinity for RNAP increased the amount of 3nt product formed as compared to the amount formed in the presence of $CarD^{WT}$ (Chapter 3, Fig. 2, (136)), demonstrating that increasing the affinity of CarD to the RNAP results in the increased stability of an RNAP-promoter complex that is able to produce the aborted transcript in this assay.

CarD regulates transcription of the majority of the *Mtb* transcriptome. While CarD's activity has been well studied at rRNA-associated promoters, the activity of CarD at other promoters throughout the genome remained elusive. To investigate the role of CarD in regulation of non-rRNA promoters, we performed RNA sequencing (RNA-seq) experiments on *Mtb* strains expressing different concentrations of the CarD protein, CarD mutants with altered affinity for RNAP, or a CarD mutant with a mutation in the DNA binding domain. Analysis of the transcriptomes of these different strains demonstrated that CarD regulates transcription from the majority of promoters in the *Mtb* chromosome, with 80% of transcripts significantly deregulated in at least one analyzed strain (Chapter 4 Fig. 1). We found that changes in the concentration of

CarD or decreases in CarD's affinity for RNAP or DNA are responsible for the largest changes in the *Mtb* transcriptome, suggesting that CarD's activity is particularly sensitive to these perturbations (Chapter 4 Fig. 3). Upregulation of many transcripts in response to loss of CarD function suggested that under normal conditions, CarD may act in part as a transcriptional repressor, possibly by hyper-stabilizing the RNAP-promoter complex (Chapter 4 Fig. 2).

CarD's regulatory activity is responsive to promoter sequence. Our analysis of the transcriptome in strains with alterations in CarD revealed large differences in the degree of deregulation between different transcripts, which suggested that CarD's activity was regulated in a transcript-dependent manner. We found that transcripts that lack a recognizable -10 promoter motif are more likely to be highly deregulated, and especially likely to be upregulated, upon alterations in CarD (Chapter 4 Fig. 4, (140)). Furthermore, we found that these promoters were more likely than promoters with canonical -10 motifs to be deregulated in multiple strains with alterations in CarD. (Chapter 4 Fig. 4) This sensitivity to changes in CarD may highlight an increased dependence on CarD to drive expression from promoters lacking a recognizable -10 motif: both -10 promoter motifs and CarD aid in formation of RP_o. Importantly, variability in CarD-mediated regulation is not completely explained by differences in -10 promoter motifs, suggesting there are more, as of yet unidentified, determinants of CarD activity.

Open Questions

The work detailed in this thesis has significantly advanced our understanding of the physiological role of CarD, the mechanism of transcriptional regulation by CarD, and the cellular regulation of the CarD protein. However, this more advanced understanding of mycobacterial transcription has raised questions about CarD's physiological role its promoter specificity.

Is CarD a regulator? Throughout this document, I have referred to CarD as a regulator of transcriptional activity and stress response. Arguably, for CarD to be a regulator of these things, CarD itself must be regulated in some way such that its activity is responsive to stimuli which would enable it to regulate the correct response to these conditions. At this point, we have not extensively investigated regulation of CarD itself. We know that transcription of carD is upregulated in response to numerous stresses but not have thoroughly investigated any corresponding changes in the protein concentration. Our RNA-seq analysis has highlighted differences in transcript abundance dependent of the concentration of CarD expressed in a cell, suggesting that alterations in CarD protein expression could effectively regulate gene expression. We have shown that CarD is post-transcriptionally regulated by the Clp protease in Mtb and Clp itself is regulated in an oxygenation dependent manner (143), which may indirectly regulate CarD activity. Furthermore, Claudio Gonzalez, our collaborator at the University of Florida, has identified a site in CarD's C-terminal domain which may be a binding site for an unidentified, allosteric regulator of CarD activity (unpublished data). All of these pieces of data are supportive of CarD's activity likely being regulated in response to environmental signals. These numerous potential avenues for post-transcriptional regulation in addition to potential transcriptional regulation lead me to think that CarD's activity is likely regulated which is why I have called it a regulator throughout this document.

What determines CarD's evolutionary conservation and essentiality? CarD is conserved in bacteria of numerous bacterial lineages with no clear evolutionary pathway to determine its retention (85–87). Additionally, it is not essential in all the bacteria that it is found in (130). Together these facts have raised interesting questions regarding CarD's evolutionary conservation. Thus far, bacterial RNAPs have only been characterized from *E. coli, Bacillus*

subtilis, Thermus thermophilus, Thermus aquaticus, and mycobacteria (109, 123, 127–129, 144). Of these organisms, *E. coli* is the only bacterium that doesn't encode a *carD* homolog and has the only RNAP that can form stable RP_o independently (31, 85–87). This may suggest that CarD is retained in organisms that form an unstable RP_o and has been lost in organisms that encode an RNAP that can form a stable RP_o independently, although more RNAP would have to be characterized to determine if this is supported.

A related question regards the essential function of CarD. While CarD is essential in mycobacteria, it is not essential in *B. subtilis*, suggesting that an RNAP that forms an unstable RP_o is not itself enough to make CarD essential (130). If CarD is essential because it is a general member of the transcriptional machinery needed for transcription of the majority of the transcriptome, then organisms with RNAPs that are unable to form a stable RP_o must encode an alternate cellular factor that can complement for loss of CarD. Alternatively, CarD's essential function could be regulation of a small subset of the transcriptome that is particularly dependent on this regulation. In this case, organisms could alleviate the essentiality of CarD by altering their genome in such a way that these relatively few essential transcripts could be adequately produced in the absence of CarD. One genetic manipulation that may free a transcript from a requirement for CarD could be gene duplication, such that the same amount of transcript could be produced from numerous less active promoters, in the absence of CarD, as is produced by one strongly activated promoter in the presence of CarD. If CarD is essential for regulation of only select transcripts, then rRNA is likely one of these transcripts. Interestingly, *B. subtilis*, which encodes a nonessential CarD, encodes 11 rRNA operons, while there is only one rRNA operon in *Mtb*, which has very stringent requirements for CarD, and two in *M. smegmatis*, in which CarD is essential but more mutations of CarD are tolerated. This supports, but does not prove, a model

in which CarD is a general member of the transcriptional machinery but its essentiality is determined by the number of rRNA operons encoded in the genome.

Why does *M. smegmatis* have more relaxed requirements for CarD as compared to *Mtb*? We consider *M. smegmatis* to have more relaxed requirements for CarD than *Mtb* because *M. smegmatis* tolerates mutations in CarD that are not viable in *Mtb* and CarD protein levels are less tightly post-transcriptionally regulated in *M. smegmatis* than in *Mtb* (101, 108). While CarD's affinity for RNAP likely determines Clp protease susceptibility in *Mtb* (136), this does not appear to be conserved in *M. smegmatis*, where mutants with different affinity for RNAP all have similar CarD concentrations (unpublished data). CarD is targeted to the Clp protease via a C-terminal epitope (133). In *M. smegmatis*, we can HA-tag the C-terminus of CarD, which should protect CarD from proteolytic degradation by Clp (133), but we have been unable to generate strains with a C-terminal tagged CarD in *Mtb*, suggesting they are not viable.

One explanation for the more relaxed requirements for CarD in *M. smegmatis* is that *M. smegmatis* encodes two rRNA operons while *Mtb* only encodes one. If regulation of rRNA is CarD's essential function, the duplication of the rRNA operon may reduce the stringency of *M. smegmatis's* requirement for CarD. Alternatively, CarD's essential function may be as a general member of the transcription machinery essential for regulation of the majority of the transcriptome. In this case, the relaxed requirements for CarD in *M. smegmatis* may reflect inherent differences between the RNAP, promoters, or other effectors of transcription in *M. smegmatis* and *Mtb*, such that *M. smegmatis* RNAP is more capable of generating stable open complexes in the absence of CarD than the *Mtb* RNAP. This second hypothesis is supported by the RNA-seq data from *M. smegmatis* in which mutations in CarD caused deregulation of a

smaller percentage of the transcriptome that similar mutations caused in *Mtb*, which suggests that CarD's full activity is required at a smaller percentage of the genome in *M. smegmatis* than *Mtb*.

What are the determinants of CarD's regulatory activity? The analysis of the *Mtb* transcriptome included in this thesis (Chapter 4) supports a role for CarD as a global regulator of transcription but highlighted pronounced differences in the degree of deregulation between strains and between transcripts within a strain. These evidence suggest that CarD's activity is responsive to characteristics of individual transcripts. Analysis of promoter motifs can partially explain this variability but is not the complete story. We additionally saw evidence supporting a role for the genomic context of genes and demonstrated that CarD's activity can be affected by at least one nucleoid associated protein. Future research is needed to determine additional factors involved in controlling CarD activity and the interplay between these factors.

How is CarD's activity affected by other transcriptional regulators? Thus far, transcription regulation by CarD has only been studied in non-stressed, exponential cultures of mycobacteria *in vivo*, and with σ^A-containing holo-RNAP *in vitro* (108, 109, 123). These studies have not addressed whether CarD can regulate transcription of RNAP containing accessory sigma factors. This is an important question with regards to CarD's function, as both CarD and accessory sigma factors are known to regulate mycobacteria's transcriptional response to stress (1, 85).

While my thesis has focused entirely on the effect of CarD on RNAP in isolation, this is not an accurate representation of transcriptional regulation in vivo, where many additional transcriptional regulators are present. Of particular interest in mycobacteria is the interplay between CarD and RbpA. RbpA is another essential transcription initiation regulator in mycobacteria that also stabilizes RPo (1, 93, 95, 98, 99). Recently, our collaborators in the

Galburt lab at Washington University have shown that CarD and RbpA can act cooperatively to stabilize RP_o *in vitro*, which suggests they may both be present at the same initiation complexes *in vivo* (In press). Future studies will address differences between the two transcriptional regulators and how their activities affect each other.

Conclusions

When I started my graduate studies, CarD was a completely novel transcription regulator that was known to associate with RNAP and regulate rRNA through an unknown mechanism (85). We now understand the mechanism by which CarD regulates transcription at rRNA promoters and comprehend the importance of different individual activities of CarD for this regulation (86, 87, 108, 109, 123). Recently, I have begun to elucidate the role of CarD in regulation on non-rRNA promoters and have identified factors that influence CarD's activity at these promoters (Chapter 4).

Based on our current understanding of CarD, I think that CarD is likely a general member of the transcription machinery in Mtb and also master regulator of gene expression that globally adjusts gene expression to a given condition. As a general member of the initiation machinery, I think CarD has evolved to stabilize open complex formation at the majority of promoters to compensate for mycobacterial RNAP's inability to inherently form stable complexes. However, I suspect that mycobacteria, particularly *Mtb*, have evolved ways to use this accessory RNAP factor to adjust gene expression. We have at least preliminary evidence that CarD's regulatory ability may be effected by transcriptional regulation, post-transcriptional stability, allosteric regulators, and the presence of nucleoid associated proteins. By adjusting any of these, and potentially other factors, Mtb may be able to globally effect gene regulation by CarD which

would allow it to tune its expression profile to a given set of conditions. In this sense, I think it is accurate to call CarD a regulator of the stress response, at least in CarD.

In *M. smegmatis*, I think CarD is definitely a general member of the transcription machinery but I am not sure that it is also a master regulator of gene expression. While CarD is also required to stabilize open complexes, I do not thing *M. smegmatis* has developed as many mechanisms to regulate CarD activity. While we do see condition specific transcriptional regulation of CarD and evidence that its activity may be affected by nucleoid associated proteins, we don't see the same post-transcriptional regulation of protein and we saw a smaller percentage of the genome deregulated by mutations in CarD. Compared the *Mtb*, *M. smegmatis* has an expanded array of sigma factors and other transcription factors which may suggest that *M. smegmatis* has evolved more CarD-independent pathways for transcriptional regulation.

While many open questions remain in regard to CarD, my studies have significantly advanced our understanding of the physiology of mycobacteria and represent a new paradigm in the field of prokaryotic transcription.

References

1. **Flentie K**, **Garner AL**, **Stallings CL**. 2016. The Mycobacterium tuberculosis transcription machinery: ready to respond to host attacks. J Bacteriol **198**:JB.00935–15–.
2. **Hershberg R**, **Lipatov M**, **Small PM**, **Sheffer H**, **Niemann S**, **Homolka S**, **Roach JC**, **Kremer K**, **Petrov DA**, **Feldman MW**, **Gagneux S**. 2008. High Functional Diversity in Mycobacterium tuberculosis Driven by Genetic Drift and Human Demography. PLoS Biol **6**:e311.
3. **Daniel TM**. 2006. The history of tuberculosis. Respir Med **100**:1862–1870.
4. **WHO**. 2015. Global Tuberculosis Report. http://www.who.int/tb/publications/global_report/en/ 1–97.
5. **Corbett EL**, **Watt CJ**, **Walker N**, **Maher D**, **Williams BG**, **Raviglione MC**, **Dye C**. 2003. The Growing Burden of Tuberculosis. Arch Intern Med **163**:1009–1021.
6. **Sudjaritruk T**, **Maleesatharn A**, **Prasitsuebsai W**, **Fong SM**, **Le NO**, **Le TT**, **Lumbiganon P**, **Kumarasamy N**, **Kurniati N**, **Hansudewechakul R**, **Yusoff NK**, **Razali KA**, **Kariminia A**, **Sohn AH**, **Sirisanthana On Behalf Of The Treat Asia Pediatric Hiv Observational Database V**. 2013. Prevalence, Characteristics, Management, and Outcome of Pulmonary Tuberculosis in HIV-Infected Children in the TREAT Asia Pediatric HIV Observational Database (TApHOD). AIDS Patient Care STDS **27**:649–656.
7. **WHO**. 2010. Treatment of tuberculosis: guidelines. 4Th Ed 160.
8. **Caminero JA**, **Sotgiu G**, **Zumla A**, **Migliori GB**. 2010. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. Lancet Infect Dis **10**:621–629.
9. **Stallings CL**, **Glickman MS**. 2010. Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. Microbes Infect.
10. **Vandal OH**, **Roberts JA**, **Odaira T**, **Schnappinger D**, **Nathan CF**, **Ehrt S**. 2009. Acid-susceptible mutants of mycobacterium tuberculosis share hypersusceptibility to cell wall and oxidative stress and to the host environment. J Bacteriol **191**:625–631.
11. **Vandal OH**, **Pierini LM**, **Schnappinger D**, **Nathan CF**, **Ehrt S**. 2008. A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. Nat Med **14**:849–854.
12. **Rengarajan J**, **Murphy E**, **Park A**, **Krone CL**, **Hett EC**, **Bloom BR**, **Glimcher LH**, **Rubin EJ**. 2008. Mycobacterium tuberculosis Rv2224c modulates innate immune responses. Proc Natl Acad Sci **105** :264–269.
13. **Ng VH**, **Cox JS**, **Sousa AO**, **MacMicking JD**, **McKinney JD**. 2004. Role of KatG catalase-peroxidase in mycobacterial pathogenisis: Countering the phagocyte oxidative burst. Mol Microbiol **52**:1291–1302.
14. **MacMicking JD**, **North RJ**, **LaCourse R**, **Mudgett JS**, **Shah SK**, **Nathan CF**. 1997. Identification of nitric oxide synthase as a protective locus against tuberculosis. Proc Natl Acad Sci **94** :5243–5248.
15. **Cooper AM**, **Segal BH**, **Frank AA**, **Holland SM**, **Orme IM**, **Orme IANM**. 2000. Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox − / − Mice Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox⁻ / ⁻ Mice **68**.
16. **Via LE**, **Lin PL**, **Ray SM**, **Carrillo J**, **Allen SS**, **Seok YE**, **Taylor K**, **Klein E**, **Manjunatha U**, **Gonzales J**, **Eun GL**, **Seung KP**, **Raleigh JA**, **Sang NC**, **McMurray**

DN, **Flynn JL**, **Barry CE**. 2008. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. Infect Immun **76**:2333–2340.
17. **Sambandamurthy VK**, **Wang X**, **Chen B**, **Russell RG**, **Derrick S**, **Collins FM**, **Morris SL**, **Jacobs WR**. 2002. A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. Nat Med **8**:1171–1174.
18. **Bange F-C**, **Brown AM**, **William R. Jacobs J**. 1996. Leucine auoxtrophy restricts growth of Mycobacterium bovis BCG in macrophages. Infect Immun **64**:1794–1799.
19. **Gordhan BG**, **Smith DA**, **Alderton H**, **McAdam RA**, **Bancroft GJ**, **Mizrahi V**. 2002. Construction and phenotypic characterization of an auxotrophic mutant of Mycobacterium tuberculosis defective in L-arginine biosynthesis. Infect Immun **70**:3080–3084.
20. **Hondalus MK**, **Bardarov S**, **Russell R**, **Chan J**, **Jacobs WR**, **Bloom BR**. 2000. Attenuation of and protection induced by a leucine auxotroph of Mycobacterium tuberculosis. Infect Immun **68**:2888–2898.
21. **Rengarajan J**, **Bloom BR**, **Rubin EJ**. 2005. Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci United States Am **102** :8327–8332.
22. **Medlab EM**, **Bernstein S**, **Sterward DS**. 1952. A Bacteriologic Study of Resected Tuberculous Lesions. Am Rev Tuberc Pulm Dis **66**:36–43.
23. **Gupta UD**, **Katoch VM**. 2005. Animal models of tuberculosis. Tuberculosis **85**:277–293.
24. **Wayne LG**, **Hayes LG**. 1996. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. Infect Immun **64**:2062–9.
25. **Deb C**, **Lee C-M**, **Dubey VS**, **Daniel J**, **Abomoelak B**, **Sirakova TD**, **Pawar S**, **Rogers L**, **Kolattukudy PE**. 2009. A novel in vitro multiple-stress dormancy model for Mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. PLoS One **4**:e6077.
26. **Sarathy J**, **Dartois V**, **Dick T**, **Gengenbacher M**. 2013. Reduced Drug Uptake in Phenotypically Resistant Nutrient-Starved Nonreplicating Mycobacterium tuberculosis. Antimicrob Agents Chemother **57**:1648–1653.
27. **Baek S-H**, **Li AH**, **Sassetti CM**. 2011. Metabolic regulation of mycobacterial growth and antibiotic sensitivity. PLoS Biol **9**:e1001065.
28. **Cunningham-Bussel A**, **Bange FC**, **Nathan CF**. 2013. Nitrite impacts the survival of Mycobacterium tuberculosis in response to isoniazid and hydrogen peroxide. Microbiologyopen **2**:901–11.
29. **Franzblau SG**, **Ann M**, **Hyun S**, **Andries K**, **Nuermberger E**, **Orme IM**, **Mdluli K**, **Angulo-barturen I**, **Dick T**, **Dartois V**, **Lenaerts AJ**. 2012. Comprehensive analysis of methods used for the evaluation of compounds against Mycobacterium tuberculosis. Tuberculosis **92**:453–488.
30. **Murakami KS**, **Darst S A**. 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13**:31–39.
31. **Saecker RM**, **Record MT**, **Dehaseth PL**. 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**:754–71.
32. **Browning DF**, **Busby SJ**. 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2**:57–65.
33. **Rojo F**. 2001. Mechanisms of transcriptional repression. Curr Opin Microbiol **4**:145–51.

34. **Lee DJ**, **Minchin SD**, **Busby SJW**. 2012. Activating transcription in bacteria. Annu Rev Microbiol **66**:125–52.
35. **Rodrigue S**, **Provvedi R**, **Jacques PÉ**, **Gaudreau L**, **Manganelli R**. 2006. The σ factors of Mycobacterium tuberculosis. FEMS Microbiol Rev **30**:926–941.
36. **Hu Y**, **Coates ARM**. 2001. Increased levels of sigJ mRNA in late stationary phase cultures of Mycobacterium tuberculosis detected by DNA array hybridisation. FEMS Microbiol Lett **202**:59–65.
37. **Manganelli R**, **Provvedi R**, **Rodrigue S**, **Beaucher J**, **Gaudreau L**, **Smith I**. 2004. Sigma factors and global gene regulation in Mycobacterium tuberculosis. J Bacteriol **186**:895–902.
38. **Manganelli R**, **Dubnau E**, **Tyagi S**, **Kramer FR**, **Smith I**. 1999. Differential expression of 10 sigma factor genes in Mycobacterium tuberculosis. Mol Microbiol **31**:715–24.
39. **Wosten M**. 1998. Eubacterial sigma-factors. FEMS Microbiol Rev **22**:127–50.
40. **Newton-Foot M**, **Gey van Pittius NC**. 2013. The complex architecture of mycobacterial promoters. Tuberculosis **93**:60–74.
41. **Bashyam MD**, **Kaushal D**, **Dasgupta SK**, **Tyagi A K**. 1996. A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. J Bacteriol **178**:4847–53.
42. **Agarwal N**, **Tyagi AK**. 2006. Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity. Nucleic Acids Res **34**:4245–57.
43. **Kremer L**, **Baulard A**, **Estaquier J**, **Content J**, **Capron A**, **Locht C**. 1995. Analysis of the Mycobacterium tuberculosis 85A antigen promoter region. J Bacteriol **177**:642–53.
44. **Bashyam MD**, **Tyagi AK**. 1998. Identification and analysis of "extended - 10" promoters from mycobacteria. J Bacteriol **180**:2568–2573.
45. **Graham JE**, **Clark-Curtiss JE**. 1999. Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). Proc Natl Acad Sci U S A **96**:11554–9.
46. **Cappelli G**, **Volpe E**, **Grassi M**, **Liseo B**, **Colizzi V**, **Mariani F**. 2006. Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: Upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. Res Microbiol **157**:445–455.
47. **Manganelli R**, **Voskuil MI**, **Schoolnik GK**, **Smith I**. 2001. The Mycobacterium tuberculosis ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. Mol Microbiol **41**:423–437.
48. **Lee J-H**, **Geiman DE**, **Bishai WR**. 2008. Role of Stress Response Sigma Factor SigG in Mycobacterium tuberculosis. J Bacteriol **190**:1128–1133.
49. **Gaudion A**, **Dawson L**, **Davis E**, **Smollett K**. 2013. Characterisation of the mycobacterium tuberculosis alternative sigma factor SigG: Its operon and regulon. Tuberculosis **93**:482–491.
50. **Raman S**, **Puyang X**, **Cheng TY**, **Young DC**, **Moody DB**, **Husson RN**. 2006. Mycobacterium tuberculosis SigM positively regulates Esx secreted protein and nonribosomal peptide synthetase genes and down regulates virulence-associated surface lipid synthesis. J Bacteriol **188**:8460–8468.
51. **Hu Y**, **Kendall S**, **Stoker NG**, **Coates ARM**. 2004. The Mycobacterium tuberculosis sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. FEMS Microbiol Lett

237:415–423.
52. **Calamita H**, **Ko C**, **Tyagi S**, **Yoshimatsu T**, **Morrison NE**, **Bishai WR**. 2005. The Mycobacterium tuberculosis SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. Cell Microbiol **7**:233–244.
53. **Ando M**, **Yoshimatsu T**, **Ko C**, **Converse PJ**, **Bishai WR**. 2003. Deletion of Mycobacterium tuberculosis Sigma Factor E Results in Delayed Time to Death with Bacterial Persistence in the Lungs of Aerosol-Infected Mice **71**:7170–7172.
54. **Kaushal D**, **Schroeder BG**, **Tyagi S**, **Yoshimatsu T**, **Scott C**, **Ko C**, **Carpenter L**, **Mehrotra J**, **Manabe YC**, **Fleischmann RD**, **Bishai WR**. 2002. Reduced immunopathology and mortality despite tissue persistence in a Mycobacterium tuberculosis mutant lacking alternative sigma factor, SigH. Proc Natl Acad Sci U S A **99**:8330–8335.
55. **Dainese E**, **Rodrigue S**, **Delogu G**, **Provvedi R**, **Laflamme L**, **Brzezinski R**, **Fadda G**, **Smith I**, **Gaudreau L**, **Palù G**, **Manganelli R**. 2006. Posttranslational regulation of Mycobacterium tuberculosis extracytoplasmic-function sigma factor sigma L and roles in virulence and in global regulation of gene expression. Infect Immun **74**:2457–61.
56. **Chen P**, **Ruiz RE**, **Li Q**, **Silver RF**, **Bishai WR**. 2000. Construction and Characterization of a Mycobacterium tuberculosis Mutant Lacking the Alternate Sigma Factor Gene, sigF. Infect Immun **68**:5575–5580.
57. **Geiman D**, **Kaushal D**, **Ko C**. 2004. Attenuation of Late -stage disease in mice infected by the Mycobacterium tuberculosis mutant lacking the SigF alternate sigma factor and identification of SigF-dependent genes by microarray analysis. Infect … **72**:1733–1745.
58. **Sun R**, **Converse PJ**, **Ko C**, **Tyagi S**, **Morrison NE**, **Bishai WR**. 2004. Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. Mol Microbiol **52**:25–38.
59. **Avarbock D**, **Avarbock A**, **Rubin H**. 2000. Differential regulation of opposing RelMtb activities by the aminoacylation state of a tRNA.ribosome.mRNA.RelMtb complex. Biochemistry **39**:11640–8.
60. **Dahl JL**, **Kraus CN**, **Boshoff HIM**, **Doan B**, **Foley K**, **Avarbock D**, **Kaplan G**, **Mizrahi V**, **Rubin H**, **Barry CE**. 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice. Proc Natl Acad Sci U S A **100**:10026–31.
61. **Primm TP**, **Andersen SJ**, **Mizrahi V**, **Avarbock D**, **Rubin H**, **Barry CE**. 2000. The Stringent Response of Mycobacterium tuberculosis Is Required for Long-Term Survival. J Bacteriol **182**:4889–4898.
62. **Klinkenberg LG**, **Lee J-H**, **Bishai WR**, **Karakousis PC**. 2010. The stringent response is required for full virulence of Mycobacterium tuberculosis in guinea pigs. J Infect Dis **202**:1397–404.
63. **Karakousis PC**, **Yoshimatsu T**, **Lamichhane G**, **Woolwine SC**, **Nuermberger EL**, **Grosset J**, **Bishai WR**. 2004. Dormancy phenotype displayed by extracellular Mycobacterium tuberculosis within artificial granulomas in mice. J Exp Med **200**:647–57.
64. **Weiss LA**, **Stallings CL**. 2013. Essential roles for Mycobacterium tuberculosis Rel beyond the production of (p)ppGpp. J Bacteriol **195**:5629–38.
65. **Ross W**, **Vrentas CE**, **Sanchez-Vazquez P**, **Gaal T**, **Gourse RL**. 2013. The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of

transcription initiation. Mol Cell **50**:420–9.
66. **Vrentas CE**, **Gaal T**, **Berkmen MB**, **Rutherford ST**, **Haugen SP**, **Vassylyev DG**, **Ross W**, **Gourse RL**. 2008. Still looking for the magic spot: the crystallographically defined binding site for ppGpp on RNA polymerase is unlikely to be responsible for rRNA transcription regulation. J Mol Biol **377**:551–64.
67. **Liu K**, **Myers AR**, **Pisithkul T**, **Claas KR**, **Satyshur KA**, **Amador-Noguez D**, **Keck JL**, **Wang JD**. 2015. Molecular Mechanism and Evolution of Guanylate Kinase Regulation by (p)ppGpp. Mol Cell **57**:735–749.
68. **Kriel A**, **Bittner AN**, **Kim SH**, **Liu K**, **Tehranchi AK**, **Zou WY**, **Rendon S**, **Chen R**, **Tu BP**, **Wang JD**. 2012. Direct regulation of GTP homeostasis by (p)ppGpp: a critical component of viability and stress resistance. Mol Cell **48**:231–41.
69. **Krásný L**, **Gourse RL**. 2004. An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. EMBO J **23**:4473–83.
70. **Tare P**, **Mallick B**, **Nagaraja V**. 2013. Co-evolution of specific amino acid in sigma 1.2 region and nucleotide base in the discriminator to act as sensors of small molecule effectors of transcription initiation in mycobacteria. Mol Microbiol **90**:569–583.
71. **Parish T**. 2014. Two-Component Regulatory Systems of Mycobacteria. Microbiol Spectr 1–14.
72. **Turkarslan S**, **Peterson EJR**, **Rustad TR**, **Minch KJ**, **Reiss DJ**, **Morrison R**, **Ma S**, **Price ND**, **Sherman DR**, **Baliga NS**. 2015. A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis. Sci Data **2**:150010.
73. **Minch KJ**, **Rustad TR**, **Peterson EJR**, **Winkler J**, **Reiss DJ**, **Ma S**, **Hickey M**, **Brabant W**, **Morrison B**, **Turkarslan S**, **Mawhinney C**, **Galagan JE**, **Price ND**, **Baliga NS**, **Sherman DR**. 2015. The DNA-binding network of Mycobacterium tuberculosis. Nat Commun **6**:5829.
74. **Baloni P**, **Chandra N**. 2015. Architectural plan of transcriptional regulation in Mycobacterium tuberculosis. Trends Microbiol **23**:123–5.
75. **Arnvig KB**, **Comas I**, **Thomson NR**, **Houghton J**, **Boshoff HI**, **Croucher NJ**, **Rose G**, **Perkins TT**, **Parkhill J**, **Dougan G**, **Young DB**. 2011. Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. PLoS Pathog **7**:e1002342.
76. **Haning K**, **Cho SH**, **Contreras LM**. 2014. Small RNAs in mycobacteria: an unfolding story. Front Cell Infect Microbiol **4**:1–11.
77. **Pelly S**, **Bishai WR**, **Lamichhane G**. 2012. A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection. Gene **500**:85–92.
78. **DiChiara JM**, **Contreras-Martinez LM**, **Livny J**, **Smith D**, **McDonough KA**, **Belfort M**. 2010. Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis. Nucleic Acids Res **38**:4067–78.
79. **Prisic S**, **Husson RN**. 2014. Mycobacterium tuberculosis Serine/Threonine Protein Kinases. Microbiol Spectr **2**.
80. **Sachdeva P**, **Misra R**, **Tyagi AK**, **Singh Y**. 2010. The sigma factors of Mycobacterium tuberculosis: Regulation of the regulators. FEBS J **277**:605–626.
81. **Gupta M**, **Sajid A**, **Sharma K**, **Ghosh S**, **Arora G**, **Singh R**, **Nagaraja V**, **Tandon V**, **Singh Y**. 2014. HupB, a nucleoid-associated protein of Mycobacterium tuberculosis, is

modified by serine/threonine protein kinases in vivo. J Bacteriol **196**:2646–57.
82. **Park ST**, **Kang C-M**, **Husson RN**. 2008. Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A **105**:13105–10.
83. **Greenstein AE**, **MacGurn JA**, **Baer CE**, **Falick AM**, **Cox JS**, **Alber T**. 2007. M. tuberculosis Ser/Thr protein kinase D phosphorylates an anti-anti-sigma factor homolog. PLoS Pathog **3**:e49.
84. **Chao JD**, **Papavinasasundaram KG**, **Zheng X**, **Chávez-Steenbock A**, **Wang X**, **Lee GQ**, **Av-Gay Y**. 2010. Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in Mycobacterium tuberculosis. J Biol Chem **285**:29239–46.
85. **Stallings CL**, **Stephanou NC**, **Chu L**, **Hochschild A**, **Nickels BE**, **Glickman MS**. 2009. CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. Cell **138**:146–59.
86. **Bae B**, **Chen J**, **Davis E**, **Leon K**, **Darst SA**, **Campbell EA**. 2015. CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. Elife **4**:1–19.
87. **Srivastava DB**, **Leon K**, **Osmundson J**, **Garner AL**, **Weiss LA**, **Westblade LF**, **Glickman MS**, **Landick R**, **Darst SA**, **Stallings CL**, **Campbell EA**. 2013. Structure and function of CarD, an essential mycobacterial transcription factor. Proc Natl Acad Sci U S A **110**:12619–24.
88. **Betts JC**, **Lukey PT**, **Robb LC**, **McAdam R a.**, **Duncan K**. 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol Microbiol **43**:717–731.
89. **Paget MS**, **Molle V**, **Cohen G**, **Aharonowitz Y**, **Buttner MJ**. 2001. Defining the disulphide stress response in Streptomyces coelicolor A3(2): identification of the sigmaR regulon. Mol Microbiol **42**:1007–20.
90. **Stewart GR**, **Wernisch L**, **Stabler R**, **Mangan J a**, **Hinds J**, **Laing KG**, **Young DB**, **Butcher PD**. 2002. Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. Microbiology **148**:3129–38.
91. **Provvedi R**, **Boldrin F**, **Falciani F**, **Palù G**, **Manganelli R**. 2009. Global transcriptional response to vancomycin in Mycobacterium tuberculosis. Microbiology **155**:1093–102.
92. **Murphy DJ**, **Brown JR**. 2007. Identification of gene targets against dormant phase Mycobacterium tuberculosis infections. BMC Infect Dis **7**:84.
93. **Bortoluzzi A**, **Muskett FW**, **Waters LC**, **Addis PW**, **Rieck B**, **Munder T**, **Schleier S**, **Forti F**, **Ghisotti D**, **Carr MD**, **O'Hare HM**. 2013. Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. J Biol Chem **288**:14438–50.
94. **Hubin EA**, **Tabib-Salazar A**, **Humphrey LJ**, **Flack JE**, **Olinares PDB**, **Darst SA**, **Campbell EA**, **Paget MS**. 2015. Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. Proc Natl Acad Sci U S A **112**:7171–6.
95. **Tabib-Salazar A**, **Liu B**, **Doughty P**, **Lewis RA**, **Ghosh S**, **Parsy M-L**, **Simpson PJ**, **O'Dwyer K**, **Matthews SJ**, **Paget MS**. 2013. The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. Nucleic Acids Res **41**:5679–91.
96. **Dey A**, **Verma AK**, **Chatterji D**. 2010. Role of an RNA polymerase interacting protein,

MsRbpA, from Mycobacterium smegmatis in phenotypic tolerance to rifampicin. Microbiology **156**:873–83.
97. **Dey A**, **Verma AK**, **Chatterji D**. 2011. Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA.Microbiology (Reading, England).
98. **Hu Y**, **Morichaud Z**, **Chen S**, **Leonetti J-P**, **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.
99. **Hu Y**, **Morichaud Z**, **Perumal AS**, **Roquet-Baneres F**, **Brodolin K**. 2014. Mycobacterium RbpA cooperates with the stress-response σB subunit of RNA polymerase in promoter DNA unwinding. Nucleic Acids Res **42**:10399–408.
100. **Verma AK**, **Chatterji D**. 2014. Dual role of MsRbpA: transcription activation and rescue of transcription from the inhibitory effect of rifampicin. Microbiology **160**:2018–29.
101. **Weiss LA**, **Harrison PG**, **Nickels BE**, **Glickman MS**, **Campbell EA**, **Darst SA**, **Stallings CL**. 2012. Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. J Bacteriol **194**:5621–31.
102. **Gulten G**, **Sacchettini JC**. 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**:1859–69.
103. **Kaur G**, **Dutta D**, **Thakur KG**. 2014. Crystal structure of Mycobacterium tuberculosis CarD, an essential RNA polymerase binding protein, reveals a quasidomain-swapped dimeric structural architecture. Proteins **82**:879–84.
104. **Gallego-Garcia A**, **Mirassou Y**, **Elias-Arnanz M**, **Padmanabhan S**, **Jiminez MA**. 2012. NMR structure note: N-terminal domain of Thermus thermophilus CdnL. J Biomol NMR **53**:355–363.
105. **Šali A**, **Blundell TL**. 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol **234**:779–815.
106. **Westblade LF**, **Campbell E a**, **Pukhrambam C**, **Padovan JC**, **Nickels BE**, **Lamour V**, **Darst S a**. 2010. Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. Nucleic Acids Res **38**:8357–69.
107. **Campbell EA**, **Muzzin O**, **Chlenov M**, **Sun JL**, **Olson CA**, **Weinman O**, **Trester-Zedlitz ML**, **Darst SA**. 2002. Structure of the bacterial RNA polymerase promoter specificity σ subunit. Mol Cell **9**:527–539.
108. **Garner AL**, **Weiss LA**, **Manzano AR**, **Galburt EA**, **Stallings CL**. 2014. CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. Mol Microbiol **93**:682–697.
109. **Rammohan J**, **Ruiz Manzano A**, **Garner AL**, **Stallings CL**, **Galburt EA**. 2015. CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. Nucleic Acids Res 1–14.
110. **Pashley CA**, **Parish T**. 2003. Efficient switching of mycobacteriophage L5-based integrating plasmids in Mycobacterium tuberculosis. FEMS Microbiol Lett **229**:211–215.
111. **Gaal T**, **Ross W**, **Estrem ST**, **Nguyen LH**, **Burgess RR**, **Gourse RL**. 2001. Promoter recognition and discrimination by EsigmaS RNA polymerase. Mol Microbiol **42**:939–54.
112. **Ross W**, **Thompson JF**, **Newlands JT**, **Gourse RL**. 1990. E. coli Fis protein activates ribosomal RNA transcription in vitro and in vivo. EMBO J **9**:3733–42.
113. **Czyz A**, **Mooney RA**, **Iaconi A**, **Landick R**. 2014. Mycobacterial RNA Polymerase

Requires a U-Tract at Intrinsic Terminators and Is Aided by NusG at Suboptimal Terminators. MBio **5**:1–10.
114. **Huff J**, **Czyz A**, **Landick R**, **Niederweis M**. 2010. Taking phage integration to the next level as a genetic tool for mycobacteria. Gene **468**:8–19.
115. **Cooper AM**, **Segal BH**, **Frank AA**, **Holland SM**, **Orme IM**, **Orme IANM**. 2000. Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox − / − Mice Transient Loss of Resistance to Pulmonary Tuberculosis. Infect Immun **68**:1231–1234.
116. **Gonzalez-y-merchand JA**, **Colston MJ**, **Cox RA**. 1996. The rRNA operons of Mycobacterium smegmatis and Mycobacterium tuberculosis: comparison of promoter elements and of neighbouring upstream genes. Microbiology **142**:667–674.
117. **Gonzalez-y-merchand JA**, **Garcia MJ**, **Gonzalez-Rico S**, **Colston MJ**, **Cox RA**. 1997. Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA. J Bacteriol **179**:6949.
118. **Masayasu N**, **Gourse R**, **Baughman G**. 1984. Regulation of the Synthesis of Ribosomes and Ribosomal Components. Annu Rev Biochem **53**:75–117.
119. **Bremer H**, **Dennis PP**. 1996. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate, p. 1553–69. *In* Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology. ASM Press.
120. **Gourse RL**, **Gaal T**, **Bartlett MS**, **Appleman J A**, **Ross W**. 1996. rRNA transcription and growth rate-dependent regulation of ribosome synthesis in Escherichia coli. Annu Rev Microbiol **50**:645–77.
121. **Chao MC**, **Rubin EJ**. 2010. Letting sleeping dos lie: does dormancy play a role in tuberculosis? Annu Rev Microbiol **64**:293–311.
122. **Gengenbacher M**, **Kaufmann SHE**. 2012. Mycobacterium tuberculosis: success through dormancy. FEMS Microbiol Rev **36**:514–32.
123. **Davis E**, **Chen J**, **Leon K**, **Darst SA**, **Campbell EA**. 2015. Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. Nucleic Acids Res **43**:433–445.
124. **Monsalve M**, **Calles B**, **Mencía M**, **Salas M**, **Rojo F**. 1997. Transcription activation or repression by phage psi 29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. Mol Cell **1**:99–107.
125. **García-Moreno D**, **Abellón-Ruiz J**, **García-Heras F**, **Murillo FJ**, **Padmanabhan S**, **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**:4586–98.
126. **Yang XF**, **Goldberg MS**, **He M**, **Xu H**, **Blevins JS**, **Norgard M V**. 2008. Differential expression of a putative CarD-like transcriptional regulator, LtpA, in Borrelia burgdorferi. Infect Immun **76**:4439–44.
127. **Whipple FW**, **Sonenshein A L**. 1992. Mechanism of initiation of transcription by Bacillus subtilis RNA polymerase at several promoters. J Mol Biol **223**:399–414.
128. **Xue Y**, **Hogan BP**, **Erie D A**. 2000. Purification and initial characterization of RNA polymerase from Thermus thermophilus strain HB8. Biochemistry **39**:14356–14362.
129. **Miropolskaya N**, **Ignatov A.**, **Bass I**, **Zhilina E**, **Pupov D**, **Kulbachinskiy a.** 2012. Distinct Functions of Regions 1.1 and 1.2 of RNA Polymerase Subunits from Escherichia coli and Thermus aquaticus in Transcription Initiation. J Biol Chem **287**:23779–23789.

130. **Kobayashia K**, **Ashikaga S**, **Aymerich S**, **Bessieres P**, **Boland F**, **Brignell SC**, **Bron S**, **Bunai K**, **Christiansen LC**, **Danchin A**, **Débarbouillé M**, **Dervyn E**, **Deuerling E**, **Devine K**, **Dreesen O**, **Errington J**, **Fillinger S**, **Foster SJ**, **Fujita Y**, **Galizzi A**, **Gardan R**, **Eschevins C**, **Fukushima T**, **Haga K**, **Harwood CR**, **Ehrlichb SD**, **Albertinid A**, **Amatid G**, **Andersene KK**, **Arnaudf M**, **Asai K**, **Ashikagah S**, **Aymerichi S**, **Kobayashi K**, **Ehrlich SD**, **Albertini A**, **Amati G**, **Andersen KK**, **Arnaud M**, **Asai K**, **Ashikaga S**, **Aymerich S**, **Bessieres P**, **Boland F**, **Brignell SC**, **Bron S**, **Bunai K**, **Chapuis J**, **Christiansen LC**, **Danchin A**, **Débarbouille M**, **Dervyn E**, **Deuerling E**, **Devine K**, **Devine SK**, **Dreesen O**, **Errington J**, **Fillinger S**, **Foster SJ**, **Fujita Y**, **Galizzi A**, **Gardan R**, **Eschevins C**, **Fukushima T**, **Haga K**, **Harwood CR**, **Hecker M**, **Hosoya D**, **Hullo MF**, **Kakeshita H**, **Karamata D**, **Kasahara Y**, **Kawamura F**, **Koga K**, **Koski P**, **Kuwana R**, **Imamura D**, **Ishimaru M**, **Ishikawa S**, **Ishio I**, **Le Coq D**, **Masson A**, **Mauël C**, **Meima R**, **Mellado RP**, **Moir A**, **Moriya S**, **Nagakawa E**, **Nanamiya H**, **Nakai S**, **Nygaard P**, **Ogura M**, **Ohanan T**, **O'Reilly M**, **O'Rourke M**, **Pragai Z**, **Pooley HM**, **Rapoport G**, **Rawlins JP**, **Rivas L a**, **Rivolta C**, **Sadaie A**, **Sadaie Y**, **Sarvas M**, **Sato T**, **Saxild HH**, **Scanlan E**, **Schumann W**, **Seegers JFML**, **Sekiguchi J**, **Sekowska A**, **Séror SJ**, **Simon M**, **Stragier P**, **Studer R**, **Takamatsu H**, **Tanaka T**, **Takeuchi M**, **Thomaides HB**, **Vagner V**, **van Dijl JM**, **Watabe K**, **Wipat A**, **Yamamoto H**, **Yamamoto M**, **Yamamoto Y**, **Yamane K**, **Yata K**, **Yoshida K**, **Yoshikawa H**, **Zuber U**, **Ogasawara N**. 2003. Essential Bacillus subtilis genes. Proc Natl Acad Sci U S A **100**:4678–83.
131. **Nickels BE**. 2009. Genetic assays to define and characterize protein-protein interactions involved in gene regulation. Methods **47**:53–62.
132. **Gulten G**, **Sacchettini JC**. 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**:1859–69.
133. **Raju RM**, **Unnikrishnan M**, **Rubin DHF**, **Krishnamoorthy V**, **Kandror O**, **Akopian TN**, **Goldberg AL**, **Rubin EJ**. 2012. Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. PLoS Pathog **8**.
134. **Schaechter M**, **Maaloe O**, **Kjeldgaard NO**. 1958. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. J Gen Microbiol **19**:592–606.
135. **Binder BJ**, **Liu YC**. 1998. Growth rate regulation of rRNA content of a marine Synechococcus (cyanobacterium) strain. Appl Environ Microbiol **64**:3346–3351.
136. **Garner AL**, **Rammohan J**, **Huynh JP**, **Onder LM**, **Chen J**, **Bae B**, **Jensen D**, **Weiss LA**, **Manzano AR**, **Darst SA**, **Campbell EA**, **Nickels BE**, **Galburt EA**, **Stallings CL**. 2016. Effects of Increasing the Affinity of CarD for RNA Polymerase on Mycobacterium tuberculosis Growth, rRNA Transcription, and Virulence (In submission).
137. **Oliveros JC**. 2007. Venny. An Interactive Tool for comparing lists with Venn Diagrams.
138. **Anders S**, **Pyl PT**, **Huber W**. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics **31**:166–169.
139. **Anders S**, **Huber W**, **S A**, **W H**. 2010. Differential expression analysis for sequence count data. Genome Biol **11**:R106.
140. **Cortes T**, **Schubert OT**, **Rose G**, **Arnvig KB**, **Comas I**, **Aebersold R**, **Young DB**. 2013. Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless

Transcriptome in Mycobacterium tuberculosis. Cell Rep **5**:1121–1131.
141. **Blasco B**, **Chen JM**, **Hartkoorn R**, **Sala C**, **Uplekar S**, **Rougemont J**, **Pojer F**, **Cole ST**. 2012. Virulence regulator EspR of mycobacterium tuberculosis is a nucleoid-associated protein. PLoS Pathog **8**.
142. **Dillon SC**, **Dorman CJ**. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat Rev Microbiol **8**:185–95.
143. **Sherrid AM**, **Rustad TR**, **Cangelosi GA**, **Sherman DR**. 2010. Characterization of a Clp Protease Gene Regulator and the Reaeration Response in Mycobacterium tuberculosis. PLoS One **5**.
144. **China A.**, **Tare P**, **Nagaraja V**. 2010. Comparison of promoter-specific events during transcription initiation in mycobacteria. Microbiology **156**:1942–1952.