

**Kinetic Characterization of Rifamycin-Resistant *M. tuberculosis* RNA  
Polymerases and Novel Therapeutic Approach for Targeting Transcription**

By

Maxwell A. Stefan

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Medicinal Chemistry)  
in the University of Michigan  
2018

Doctoral Committee:

Professor George A. Garcia, Chair  
Professor Amanda Garner  
Professor Nouri Neamati  
Professor Ronald Woodard

“insert your own witty quote here.”

- Me

**Maxwell A. Stefan**

**mastefan@umich.edu**

**ORCID iD: 0000-0001-8196-2452**

**© Maxwell A. Stefan 2018**

**In memory of Shelley Elizabeth Hutsler**

**1953-2012**



## **ACKNOWLEDGEMENTS**

During my time here at the University of Michigan College of Pharmacy there have been several people who have helped me grow as both a scientist and as a person. The foremost is my boss Dr. George Garcia, I would like to thank you for all your support over these last several years. I have learned tremendously from your mentorship, together I feel we have made some really great strides in the lab. I want to thank you for supporting me throughout my tenure at UM and for allowing me the freedom to pursue my scientific interests (many PI's would not). Thank you for everything. I would also like to thank the members of my committee Dr. Ron Woodard, Dr. Nouri Neamati, and Dr. Amanda Garner for your guidance and support throughout the course of my dissertation research.

I would like to thank the members of the Garcia Lab, both past and present. Dr. Irosha Nawarathne was my first encounter in my laboratory experience at the UM and helped me to "learn the ropes" in the lab, thank you for your mentorship. I would also like to thank Dr. Anthony Emanuele for dealing with my borderline psychotic behavior at times and also science. Dr. Nathan Scarf for his work on the malachite green assay and helping me to pay the mortgage for a time. Nicholas Ragazzone for our weekly gossip and coffee. Fatima Ugur for being one of the best students I have mentored, I guess you don't know what you got till it's gone and of course both Katie Guild and Shireen Askar.

There are also several people who made Michigan feel more like home, Jen, Michael, Eric, Erin, Brandt, Tony, Helen, Taylor, and Joe (no last names needed), you all are a hoot and I'm glad we all got to do this together. The world isn't big enough for us to not run into each other again. I would also like to thank all of my Michigan family, both the McKinnon and Sasena clans, thank you for being there for me and inviting me to all of your Thanksgivings and Easters and a special shout out to Uncle Dick and Aunt Margaret for not letting me starve by having me for Sunday dinners.

Lastly, I would like to thank my family. Even though my mom, Shelley, could not be here to see me finish, I know I would not be here without her support, for that I am eternally grateful. I would like to thank by dad, Frank, for all of his support and my sister Elizabeth.

## TABLE OF CONTENTS

DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF FIGURES.....	vii
LIST OF TABLES.....	x
ABSTRACT.....	xi
CHAPTER	
<b>I. Introduction.....</b>	<b>1</b>
Antibiotic Resistance.....	1
Tuberculosis and <i>Mycobacterium tuberculosis</i> .....	2
<i>M. tuberculosis</i> Pathogenesis.....	4
Treatments for tuberculosis and Drug Resistance.....	6
Rifampin Resistance .....	15
Fitness Defects and Compensatory Mutations.....	19
RNA polymerase Structure and Function.....	22
Inhibitors of Bacterial RNA polymerase.....	25
Mycobacterial Transcription Regulator CarD.....	30
Research Objectives.....	32
References.....	34
<b>II. Purification and Characterization of MTB RNA polymerase.....</b>	<b>48</b>
Materials and Methods.....	50

Results.....	57
Discussion.....	69
Notes to Chapter II.....	75
References.....	76
Appendix.....	80
<b>III. <i>In vitro</i> Characterization of Rifampin Resistant <i>M. tuberculosis</i> RNA polymerases.....</b>	<b>87</b>
Materials and Methods.....	89
Results.....	95
Discussion.....	103
Notes to Chapter III.....	110
References.....	111
Appendix.....	116
<b>IV. Development of a Fluorescence Polarization Assay to Identify Inhibitors of MTB Transcriptional Regulator CarD .....</b>	<b>122</b>
Materials and Methods.....	123
Results.....	131
Discussion.....	146
Notes to Chapter IV.....	152
References.....	153
Appendix.....	156
<b>V. Concluding Summary.....</b>	<b>171</b>
References.....	178

## LIST OF FIGURES

### Figure

I-1.	Tuberculosis health facts. ....	3
I-2.	Composition of the Mycobacterium tuberculosis cell wall.....	10
I-3.	Pipeline of drugs for the treatment of tuberculosis.....	15
I-4.	Sequence alignment spanning the RRDRs of the <i>E. coli</i> , <i>T. thermophilus</i> and MTB $\beta$ -subunits of RNAP.....	17
I-5.	Compensatory mutation sites.....	20
I-6.	Structure of RNAP and kinetics of transcription.....	24
I-7.	Inhibitors of RNA polymerase and functional region of RNAP which they target.....	27
I-8.	Interaction of CarD with RNAP.....	31
II-1.	Overexpression of MTB RNAP with pMTBRP.....	57
II-2.	Purification of MTB RNAP with 1st-generation Duet Expression System..	58
II-3.	Purification of MTB RNAP with 2nd-generation Duet Expression System..	60
II-4.	Purification of MTB RNAP with 1st-generation Duet Expression System..	61
II-5.	Determination of percent active complex via primer extension assay .....	62
II-6.	Inhibition of WT and RMPR MTB RNAPs by rifampin.....	64
II-7.	Structural basis for inhibition of rifampin by RMPR mutant <i>E. coli</i> RNAPs..	67
III-1.	Open-promoter complex studies with MTB RNAP in the presence and absence of CarD.....	96

III-2.	Elongation Rate of WT and RMPR MTB RNAPs.....	98
III-3.	Termination studies at the <i>tuf</i> and <i>metK</i> terminators with all MTB RNAPs included in this study in the presence and absence of NusA and NusG...	100
III-4.	Percent hydrolysis of WT and RMPR MTB RNAPs.....	102
III-5.	Structure of <i>Mycobacterium smegmatis</i> RNAP with full transcription bubble.....	105
III-6.	Concluding observations from kinetic studies of MTB RNAPs.....	108
IV-1.	Characterization of labeling sites on CarD.....	132
IV-2.	Optimization of CarD FP Assay.....	134
IV-3.	CarD affinities for DNA, RNAP, and the open-promoter complex (RPo).	135
IV-4.	Competition experiments with unlabeled CarD variants.....	137
IV-5.	Fidaxomicin inhibition of CarD association with RPo.....	138
IV-6.	Pilot screen run with CarD FP assay with AnalytiCon Library.....	139
IV-7.	Maybridge 24K HTS with the CarD FP assay.....	140
IV-8.	Schematic of criteria for selection of hits from the primary screen to compound ordering.....	142
IV-9.	Monitoring direct binding of small molecules to CarD and eIF4E by SPR.	145
V-1.	Model of MAS-16 docked into MTB CarD using MOE.....	177

## Appendix Figure

II-A1. Schematic showing the evolution of the purification of MTB RNAP .....	81
II-A2. Purification of MTB RNAP with cation-exchange chromatography.....	82
II-A3. Purification of MTB RNAP with immobilized-metal affinity chromatography.....	83
II-A4. Purification of MTB RNAP with anion-exchange chromatography.....	84
II-A5. Determination of percent of contaminating endogenous WT E. coli RNAP in the RNAP purifications.....	85
II-A6. Rifampin IC50 curves for WT and RMPR E. coli and MTB RNAPs.....	86
III-A1. Sequence of source DNA template used in Chapter III.....	118
III-A2. SDS-PAGE of all proteins used in Chapter III.....	118
III-A3. Gel images and plots for determined open-promoter complex half-life...	119
III-A4. Plots used to determine the elongation rate for MTB RNAPs.....	120
III-A5. Effect of NusA and NusG on elongation rate with WT MTB RNAP.....	121
IV-A1. LCMS Characterization of WT CarD.....	158
IV-A2. LCMS Characterization of CarD R47E.....	159
IV-A3. LCMS Characterization of CarD K90A.....	160
IV-A4. LCMS Characterization of CarD T8C BODIPY.....	161
IV-A5. LCMS Characterization of CarD T24C BODIPY.....	162
IV-A6. LCMS Characterization of CarD T26C BODIPY.....	163
IV-A7. LCMS Characterization of CarD D68C BODIPY.....	164
IV-A8. LCMS Characterization of CarD T152C BODIPY.....	165
IV-A9. LCMS Characterization of CarD S162C BODIPY.....	166
IV-A10. CarD FP Assay at 25°C.....	167
IV-A11. Titration of rrnAP3 DNA in CarD FP assay.....	167
IV-A12. Retest plate data for confirmation from MB24K primary screen.....	168
IV-A13. Curves used to generate IC50 values for hits from the MB24K HTS.....	170

## LIST OF TABLES

### Table

I-1.	First line treatments for Tuberculosis.....	8
I-2.	Second line treatments for Tuberculosis.....	12
IV-1.	Reconfirmed hits from fresh powders for the CarD MB24K HTS.....	143

### Appendix Table

II-A1.	Primers used in Chapter II.....	80
III-A1.	Primers used in Chapter III.....	116
III-A2.	Plasmids used in Chapter III.....	117
III-A3.	Percent of active RNAP from each preparation as determined from primer extension.....	117
IV-A1.	Primers used in Chapter IV.....	156
IV-A2.	Characterization of mutant and BODIPY-labeled CarD variants.....	157
IV-A3.	Physiochemical properties of reconfirmed hits from MB24K HTS.....	169



## ABSTRACT

Tuberculosis (TB) remains a critical threat to global human health. In 2016, 1.7 million people died from the disease. Rifampin (RMP) remains a key component of the front-line treatment for TB, though resistance has presented challenges for its efficacy. Resistance to RMP (RMP<sup>R</sup>) primarily occurs through point mutations of its target, RNA polymerase, within the rifamycin resistance determining region (RRDR) in the  $\beta$ -subunit. Three mutations constitute the bulk of RMP<sup>R</sup>,  $\beta$ D435V,  $\beta$ H445Y, and  $\beta$ S450L, with the latter being most prevalent in clinically resistant isolates. The molecular mechanisms which yield the observed distribution of RMP<sup>R</sup> mutations in MTB have been speculated upon; however, detailed *in vitro* studies of *Mycobacterium tuberculosis* (MTB) RNAP to elucidate those mechanisms have been lacking. This has likely been due, in part, to difficulty in acquiring pure MTB RNAP. To surmount this, an optimized methodology for the expression and purification of highly pure and active MTB RNAP is described. Co-expression of multiple vectors harboring all subunits of the RNAP holoenzyme allows for *in vivo* assembly of the holo RNAP complex. An optimized purification method was developed to acquire stoichiometric holo RNAP with high activity.

*In vivo* fitness defects have been observed in RMP<sup>R</sup> mutants of MTB RNAP. These defects have been found to be ameliorated by the presence of secondary mutations in double-psi  $\beta$ -barrel (DPBB) of the RNAP  $\beta'$ -subunit. To identify factors contributing to this fitness defect, several *in vitro* transcription assays were utilized to probe initiation, elongation, termination and RNA primer hydrolysis with the wild-type and RMP<sup>R</sup> RNAPs. Secondary, compensatory mutations are predominantly associated with the  $\beta$ S450L mutant, therefore this mutant was also studied in the presence of secondary mutations. We found that the RMP<sup>R</sup> mutants exhibit significantly poorer termination efficiency relative to wild-type, an important factor for proper gene expression. This may contribute to the relative prevalence of the RMP<sup>R</sup> mutants observed in MTB clinical isolates. We also found that several mechanistic aspects of transcription of the Rif<sup>R</sup> mutant RNAPs are impacted relative to wild-type, particularly the stability of the open-promoter complex and elongation rate. For the  $\beta$ S450L mutant, these defects are mitigated in the presence of secondary mutations in the DPBB of the  $\beta'$ -subunit, making the intrinsic properties of this mutant similar to those of the wild-type. These data provide insight into the cost of antibiotic resistance to the fitness of the organism and a mechanistic basis for how MTB alleviates fitness defects associated with drug resistance.

Drug resistant TB has become pervasive in large part due to a lack of novel therapeutics which act by new mechanisms of action. CarD is a global transcription regulator which acts by stabilizing the open-promoter complex of MTB RNAP and has been shown to be required for MTB viability. This suggests that CarD may be an effective and novel target for therapeutic discovery for the treatment of tuberculosis. A fluorescence polarization assay which monitors the association of MTB RNAP, native

rRNA promoter DNA and Bodipy-CarD has been developed, optimized and validated. A high throughput screen has been conducted to identify and characterize small molecule inhibitors which block the CarD•RNAP•DNA interaction. Several preliminary hits have been identified from this screen and initial secondary characterizations have been performed. This project will be the foundation for further investigation of CarD's potential as a therapeutic target.

## CHAPTER I

### Introduction

#### *Antibiotic Resistance*

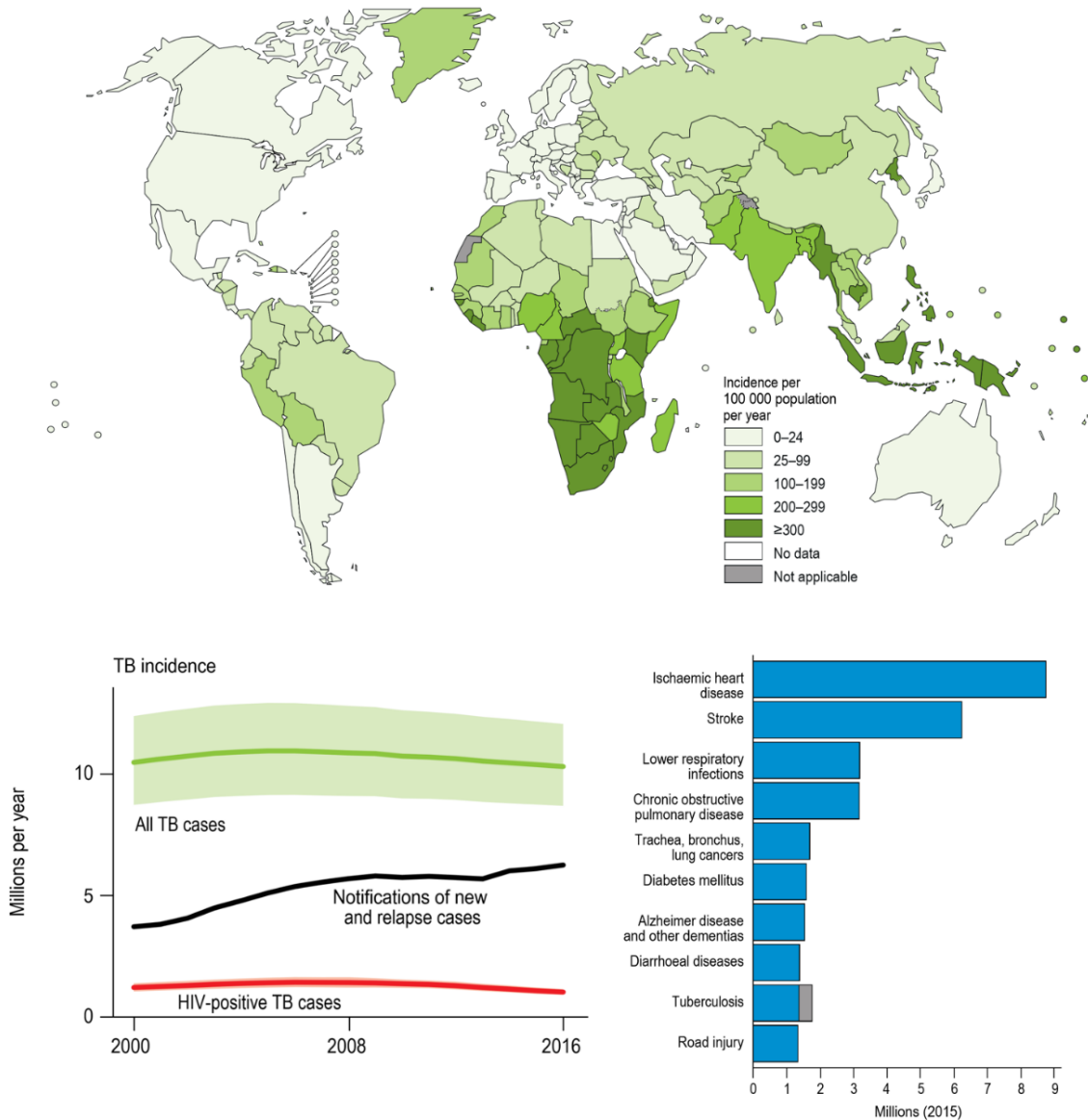
The discovery of antibiotics revolutionized the treatment of many common illnesses and infections which were almost certainly a death sentence previously. The mid 1900's saw the discovery of several novel antibiotic classes and is known as the "golden era" for antibiotics.<sup>1</sup> This "golden era" was short-lived, by the mid-1980's rising levels of resistance to established treatments and a void due to the lack of new antibiotics in the pipeline has led to a new health crisis.<sup>2</sup> Clinical resistance has been observed in all of the roughly dozen classes of antibiotics in use today. Most antibiotic resistance is the result of pervasive misuse of antibiotics in many sectors, from personal medicine to farming and agriculture.<sup>2</sup> Resistance to antibiotics occurs by several mechanisms including antibiotic efflux, enzymatic modification of the drug or the drug target.<sup>3</sup> Every year in the United States alone, 2 million people are infected with an antibiotic resistant pathogen and 23,000 deaths are attributed to antibiotic resistance.<sup>4</sup> Identification of new treatments has been tenuous, and with the withdrawal of many pharmaceutical companies from infectious disease therapeutics research, discovery of new antibiotics is lacking. To compound the issue, one in five new drug candidates which reach Phase I clinical trials ultimately fail.<sup>5</sup>

## *Tuberculosis and Mycobacterium tuberculosis*

Tuberculosis (TB) has plagued mankind for thousands of years and remains a major global health threat to this day. TB is the leading cause of death from infectious disease and ninth leading cause of mortality worldwide, with 1.7 million deaths attributed to the disease in 2016 (including deaths of 400,000 individuals co-infected with HIV/AIDS).<sup>6</sup> The cause of TB is the pathogen *Mycobacterium tuberculosis* (MTB). An estimated 2 billion people or one-third of the global population are infected with MTB and in 2016 there were 10.4 million new cases of the disease.<sup>6</sup>

A vast majority of the 2 billion people infected with MTB are asymptomatic; they carry the pathogen in a latent state.<sup>6</sup> This latent tuberculosis infection (LTBI) occurs in 90% of TB cases. Roughly 5% of TB cases result in active infection in the initial 18 months post exposure, after which people with LTBI have a 5% risk of reactivation during their lifetime.<sup>7-8</sup> People with a compromised immune system, such as those co-infected with HIV/AIDS, have an increased risk of reactivation. Additionally, individuals with diabetes and those who consume alcohol or smoke are more likely to contract TB.<sup>6</sup> For individuals with LTBI the use of either the tuberculin skin test or interferon gamma (IFN- $\gamma$ ) release assays, which indicate the presence of MTB antigens is used for diagnosis.<sup>6, 9</sup> Four diagnostic methods are utilized for confirming the presence of active TB: X-ray, rapid molecular tests (Xpert MTB/RIF Assay), sputum smear microscopy, and other culture based methods. Rapid molecular tests (RMT) such as the Xpert MTB/RIF assay is the preferred diagnostic tool by the World Health Organization (WHO) as it can detect the presence of MTB as well as resistance to rifampin (the first-in-line treatment for TB) within a few hours.<sup>6</sup> Although RMT is the preferred diagnostic method, culture-based methods

remain a requisite tool for monitoring disease progression and development of secondary resistance



**Figure I-1: Tuberculosis health facts.** A) Map of estimated global TB incidence rate for 2016. B) Graphical representation of TB incidence with all TB cases (green), notification of new and relapse cases (black), and HIV-positive TB cases. C) Top causes of death worldwide in 2015 (grey bar represents comorbidity with HIV).<sup>6</sup>

### *M. tuberculosis pathogenesis*

MTB is communicable disease, which spreads when aerosolized pathogen is inhaled. Inhaled bacteria deposit in the distal alveoli and are internalized by macrophages and other phagocytic cells. Uptake of MTB into phagocytic cells is mediated by several cellular receptors including C-type lectin, scavenger receptors, and complement receptors.<sup>10</sup> It is currently thought that route of pathogen entry into cells by different receptors plays a critical role in bacterial pathogenesis.<sup>11-12</sup>

Even though MTB has been shown to colonize many cell types, a majority of research has focused on macrophages. Interestingly transcriptome data of MTB occupying different cell types varies possibly indicating multiple routes of pathogenesis.<sup>11</sup> Once the bacteria are phagocytized by the macrophage, MTB prevents maturation of the phagosome, avoiding formation of a degradative environment within the compartment.<sup>13-15</sup> There is growing evidence that MTB can replicate and persist in the cytosol.<sup>16</sup> This escape from the phagosome compartment into the cytosol is thought to be mediated by Esx-1, a type VII secretion system (TSSS), which weakens the membrane of the phagosome.<sup>17</sup> Autophagy, the endogenous mechanism to degrade cellular components and debris, is activated by IFN- $\gamma$  and vitamin D, plays a role in the cellular recognition and targeting of MTB for clearance.<sup>18-19</sup> Intracellular survival of MTB in macrophages is required for persistence of MTB. Activation of macrophages occurs from many stimuli including IFN- $\gamma$ , vitamin D, and tumor necrosis factor with the former being produced by T lymphocytes.<sup>18, 20</sup> Cell to cell spread of MTB is thought to occur by necrotic release of bacilli into the extracellular matrix.

Persister bacteria are isolated by the immune response within hardened structures called granulomas. Granulomas are composed of macrophages, lymphocytes and other immune cells and function to isolate MTB from nutrients, oxygen, and induce oxidative stress to produce a bactericidal environment.<sup>21</sup> There are several granuloma subtypes including: caseous, non-necrotizing, neutrophil-rich, mineralized, and completely fibrotic or cavitory.<sup>22</sup> It is this heterogeneity which produces individual and isolated microenvironments which dictate bacteria pathology. Lesions which compromise granuloma structure can release the bacteria, leading to dissemination and reactivation of TB infection.

In order to survive within granulomas, MTB must enter a low metabolic state leading to a LTBI.<sup>23</sup> Part of this transition is mediated by the stringent response.<sup>24</sup> The stringent response is initiated by a drastic change in environmental conditions, mainly amino acid starvation; however, other cues have been shown to elicit the same response.<sup>24</sup> The stringent response reprograms gene expression the bacteria adjust to its new environment. Part of this differential expression, at least in *E. coli*, is the direct modulation of RNA polymerase resulting in the reduction of rRNA and tRNA transcription.<sup>25</sup> Other changes in transcription include an increase in gene expression associated with amino acid transport and biosynthesis. Guanosine 5'-diphosphate-3'diphosphate (ppGpp) is a key effector molecule that mediates this reprogramming in MTB.<sup>24-25</sup> Additional environmental cues induce the transition to the latent phase such as hypoxia.



### *Treatments for tuberculosis and Drug Resistance*

Tuberculosis has been pervasive for several millennia, having been recorded in many ancient geographically diverse cultures including the Egyptians, Peruvians, Chinese and Greek. Early recommendations for treatment of TB included fresh air, milk, and maritime travel by Greek physicians.<sup>26</sup> In the late 1800's Nobel laureate Herman Heinrich Robert Koch derived an extract from live MTB which he referred to tuberculin. He injected himself with the substance and it was later used to treat cattle by Danish veterinarians.<sup>27</sup> What Koch did not know at the time was that he had not found a cure for TB but had rather discovered the MTB antigens which are still used to this day for the diagnosis of MTB infections. Later methodologies for the treatment of TB were mostly geared toward isolation of the patients in sanatoriums, preventing the spread of the disease.

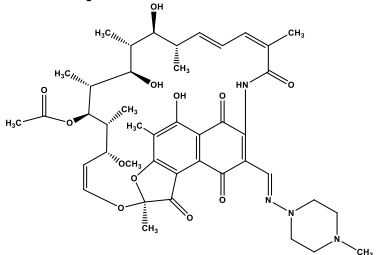
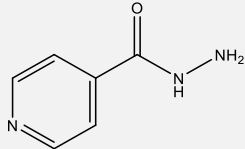
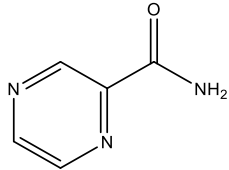
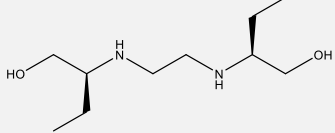
In the early 1900's Albert Calmette and Camille Guérin began pioneering the first attempt at a vaccine for TB. Their work led to the discovery of the BCG (Bacille Calmette-Guérin) vaccine in 1921.<sup>28</sup> This vaccine was used heavily after World War II by the World Health Organization (WHO) in an effort to eliminate the high prevalence of TB across the European continent and is widely accepted as one of the first major campaigns of mass vaccination.<sup>29</sup> Over 14 million people were inoculated with the BCG vaccine. Though this vaccine was heavily used during the mid-20<sup>th</sup> century, the WHO now discourages its use with the exception of those predisposed to exposure and infants, because of a lack of uniform efficacy as well as confounding tuberculin skin testing.<sup>6, 30</sup>

Treatment of TB was revolutionized by the advent of chemotherapies in the mid 1900's. The discovery and use of para-amino salicylic acid and thiosemicarbazone

represented the first chemotherapies used to treat TB infection and streptomycin was the first to show bactericidal effects on MTB.<sup>31</sup> Major strides in curbing mass TB infection prevalence were not achieved until the discovery of isoniazid in 1952 and rifamycins in 1957.<sup>32-33</sup> These two therapies radically changed the treatment of TB infections and are currently the mainstay of first line TB treatment.

Current recommendations for the treatment of TB involve a regime of multiple antibiotics. For new TB infections the treatment regime includes 2 months of rifampin, isoniazid, pyrazinamide, and ethambutol in a period known as the induction phase (see **Table I-1**).<sup>6, 34</sup> Following this initial phase pyrazinamide and ethambutol are discontinued and treatment continues for an additional 4 month period which is known as the consolidation phase.<sup>34</sup> Together these treatments represent the cornerstone for the treatment of drug susceptible MTB infections used clinically. Although the regime described above shows efficacy against drug susceptible MTB, drug resistance to all of the first line treatments have been reported (see **Table I-1**).

**First Line Antibiotics**

Drug	Target/ Mechanism of Action	Resistance gene
<p><b>Rifampin</b></p> 	<p>RNA polymerase (<math>\beta</math>-subunit) Inhibition of target enzyme</p>	<p><i>rpoB</i></p>
<p><b>Isoniazid</b></p> 	<p>FASII enoyl-ACP reductase Inhibition of target enzyme</p>	<p><i>inhA, katG</i></p>
<p><b>Pyrazinamide</b></p> 	<p>Target unknown Disruption of intracellular pH</p>	<p><i>pncA, pncA promoter</i></p>
<p><b>Ethambutol</b></p> 	<p>Arabinosyl transferase Inhibits cell wall biosynthesis</p>	<p><i>embB</i></p>

**Table I-1: First line treatments for Tuberculosis**

Rifampin (RMP) is arguably the most important component of the current treatment for TB (see **Table I-1**). RMP is derived from the ansamycin class of antibiotics discovered in extracts from *Amycolatopsis rifamycinica* found in soil samples from the pine forests of the French Riviera.<sup>32</sup> RMP, a semisynthetic rifamycin derivative which targets the  $\beta$ -subunit of RNA polymerase, was a groundbreaking addition to the arsenal of TB treatments because it shortened the treatment time for TB infections from 18 months to 9 months.<sup>35</sup> A critical property of rifampin is its sterilization ability which allows it to kill latent MTB.<sup>35</sup> The exact mechanisms of rifampin function and resistance will be described in detail in the next section.

Another critical component in the treatment regime is isoniazid which inhibits mycolic acid biosynthesis, a major component of the mycobacterial outer membrane (MOM) section of the MTB cell wall (see **Figure I-2**).<sup>36</sup> Isoniazid is a prodrug which enters the cell via passive diffusion. Its prodrug form is activated by KatG a catalase-peroxidase which leads to the formation of reactive oxygen species, superoxide and the isonicotinyl radical.<sup>37</sup> The former causes cell damage to several macromolecular cell components and the isonicotinyl radical covalently modifies NAD<sup>+</sup> which leads to the inhibition of FASII enoyl-ACP reductase (InhA). This compound has bactericidal effects against actively replicating MTB; however, it shows limited efficacy against LTBI.<sup>37</sup> Primary resistance to isoniazid arises from point mutations in either the *katG* gene, preventing activation of prodrug, or the target *inhA* gene, preventing interaction of covalently modified NAD interaction with InhA.<sup>37-38</sup> (see **Table I-1**).

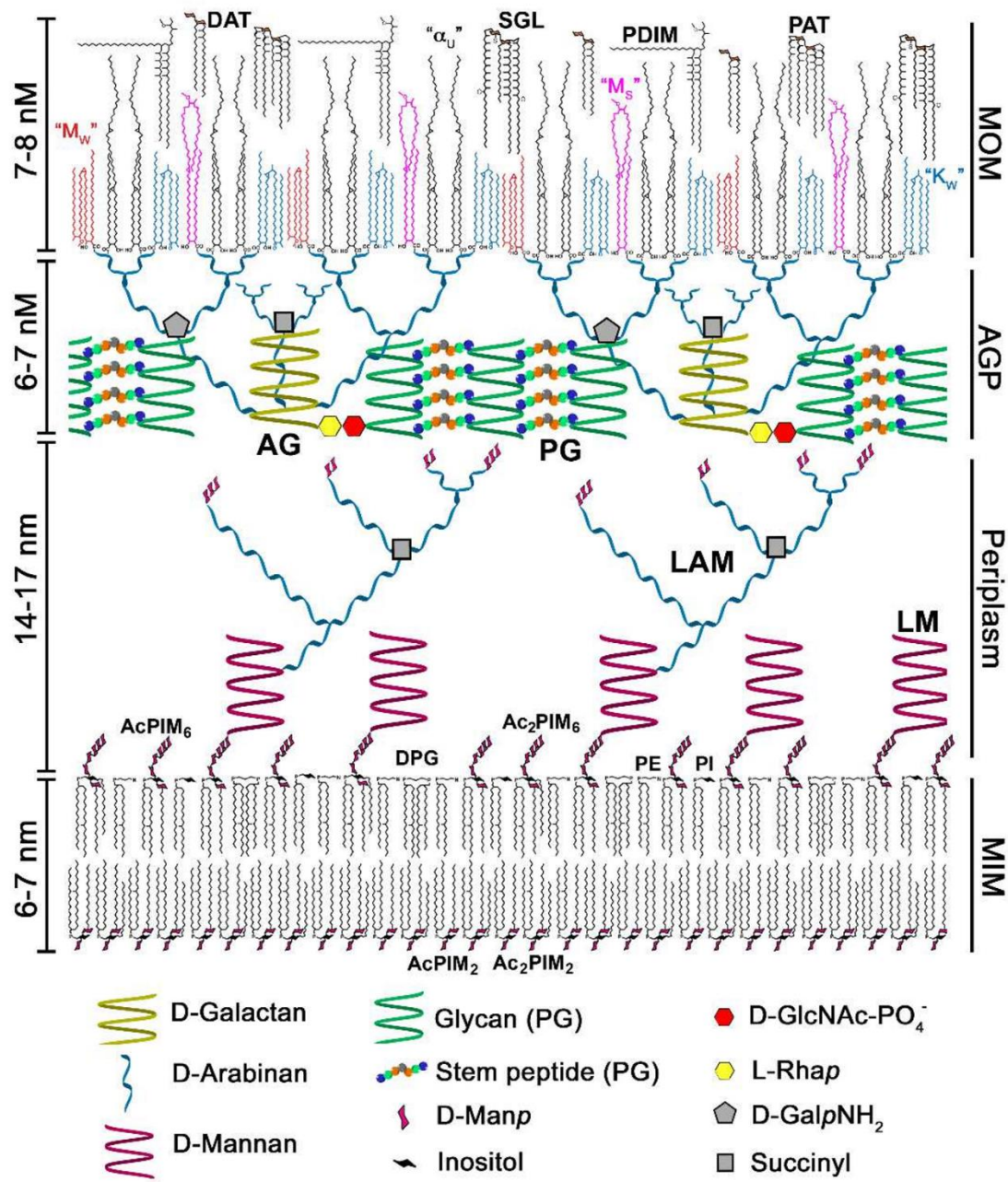


Figure I-2: Composition of the *Mycobacterium tuberculosis* cell wall.<sup>39</sup>

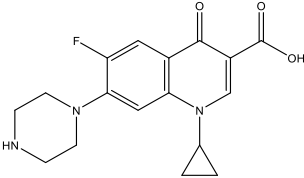
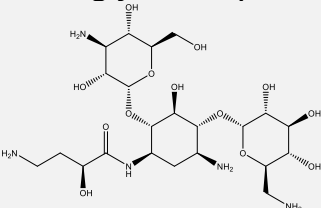
The exact mechanism of action for pyrazinamide still remains unclear; however, its specificity for killing of mycobacteria within the *Mycobacterium tuberculosis* complex (*M. tuberculosis*, *M. africanum*, and *M. microti*) and sterilizing capabilities were critical for reduction of the time of treatment from 9 months to 6 months in conjunction with RMP.<sup>40</sup> Like isoniazid, pyrazinamide is a prodrug which is converted to its active form, pyrazinoic acid, by a bacterial amidase nicotinamidase/pyrazinamidase or Pzase (*pncA*). In its anionic form, pyrazinamide has no antibiotic effects; however, it is actively effluxed and protonated in the extracellular matrix, upon reentry into the cell it causes an acidification of the cell which leads to cell death.<sup>40</sup> Primary resistance to pyrazinamide comes from modification of *pncA*, either from point mutations in the gene or via deletions and/or insertions (see **Table I-1**). Another reported mechanism of resistance has been attributed to modification of the *pncA* promoter, likely altering expression of *pncA*.<sup>41</sup>

Lastly ethambutol is included during the induction phase, mostly as a defense against resistance to the other three drugs mentioned above. The exact mechanism of action for ethambutol is still unclear but it has been shown to inhibit arabinosyl transferase, an enzyme which forms the arabinogalactan component of the MTB cell wall (see **Figure I-2**).<sup>42</sup> Its use is limited in children due to toxicity and this drug is associated with optic neuropathy further complicating the use in the induction phase.<sup>43</sup> Resistance to ethambutol occurs through mutations in the *embCAB* operon which codes for arabinosyl transferase. The vast majority of these mutations occur within codons 306-497 of *embB*, with the most frequent occurring at codon 306 (see **Table I-1**).<sup>44</sup>

In 2016, there were 600,000 cases of rifampin resistant (RMP<sup>R</sup>) cases of MTB reported of which, 490,000 were classified as multiple-drug resistant TB cases (MDR-

TB).<sup>6</sup> MDR-TB is defined as MTB which is resistant to both rifampin and isoniazid. A majority of these cases occurred in India, China, and the Russian Federation. Development of MDR-TB substantially changes treatment time, regimen, and expected outcome.<sup>6</sup> The WHO recommends a treatment regime lasting at least 20 months for MDR-TB.<sup>34</sup> Reported success rates for individuals with MDR-TB is 50% which is significantly lower than the roughly 80% success rate for those with drug susceptible TB.<sup>6</sup> Direct cost of treatment of TB infection in the USA also increases from on average \$17,000 to \$150,000 with MDR-TB infections (though this number is significantly lower in developing countries).<sup>45</sup> Several second-line treatments are recommended for individuals diagnosed with MDR-TB, these include a fluoroquinolone and at least one injectable aminoglycoside such as amikacin (see **Table I-2**).<sup>6, 34, 46</sup>

### Second Line Antibiotics

Drug	Target/ Mechanism of Action	Resistance gene
<b>Fluoroquinolone (Ciprofloxacin)</b> 	DNA gyrase and topoisomerase IV Inhibition of Target Enzyme	<i>gyrA, gyrB</i>
<b>Aminoglycoside (amikacin)</b> 	30S Ribosome subunit Disrupts Protein Synthesis	<i>rpsL, eis, rrs</i>

**Table I-2: Second line treatments for Tuberculosis**

A fluoroquinolone such as ciprofloxacin, gatifloxacin, levofloxacin or levofloxacin, are recommended for MDR-TB treatment regimes. Fluoroquinolones target bacterial DNA replication, more specifically DNA gyrase and topoisomerase IV.<sup>47</sup> An addition to a

fluoroquinolone an injectable aminoglycoside is administered.<sup>34</sup> Examples from this antibacterial class include amikacin, capreomycin, kanamycin, and streptomycin. These antibiotics function by disrupting ribosome function. Addition of these treatments to the patient's regimen can be difficult because many can cause undesirable adverse effects which can lead to poor patient compliance.<sup>9</sup> Resistance to both of these drug types have been reported. Primary resistance to fluoroquinolones occurs from modification to the *gyrA* and *gyrB* genes encoding DNA gyrase and topoisomerase IV.<sup>48</sup> Though the mechanisms for each of the aminoglycosides differ slightly, resistance primarily occurs through target modification of S12 protein (*rpsL*), aminoglycoside acetyltransferase (*eis*) or via an inactivating mutation in the 16S rRNA (*rrs*) (see **Table I-2**).<sup>49</sup> Other second line treatments for MDR-TB include ethionamide, D-cycloserine, para-aminosalicylic acid.<sup>9</sup>

Clinical isolates which present resistance to all of the antibiotics mentioned up to this point are characterized as extensively drug resistant TB (XDR-TB).<sup>6</sup> People who have this form of TB have a 30% positive outcome (28% die, 21% fail treatment, and 20% are lost to follow up). It is estimated that 6.2% of MDR-TB cases are actually XDR-TB and this form of TB has been observed in 123 countries with a majority of cases found in Ukraine and the Russian Federation.<sup>6</sup> Cost for treatment of XDR-TB in the USA skyrockets from \$150,000 for MDR-TB to nearly \$500,000 for XDR-TB.<sup>45</sup> Needless to say, increasing prevalence of XDR-TB, low efficacy of the current treatment regimen and lack of novel treatments represents a critical global health threat.

As a result of the recognition of the scope of the TB pandemic as a global threat, there has been some progress in identifying new antitubercular treatments. In 2012, the FDA approved the first new drug for TB in over 50 years, bedaquiline (BDQ), for use in

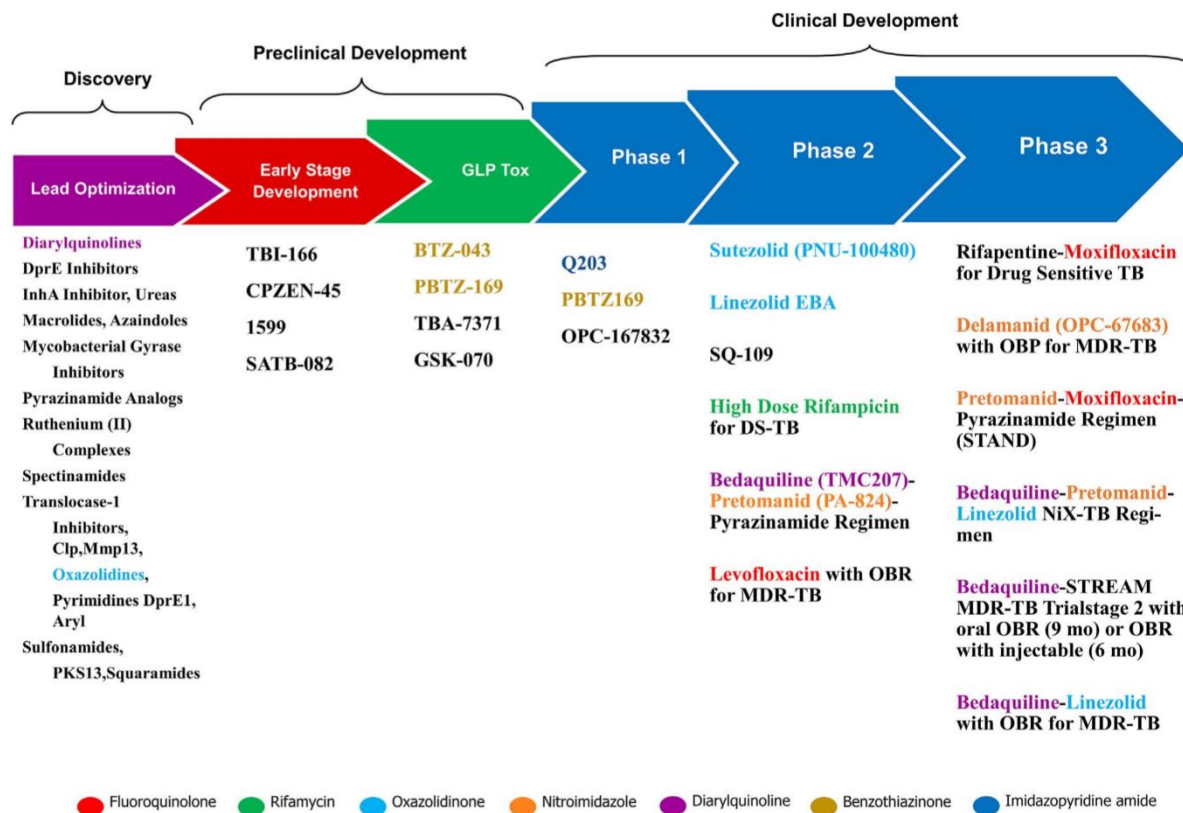


the treatment of MDR-TB.<sup>34, 50-51</sup> Bedaquiline is a diarylquinoline which targets subunit C of ATP synthase (*atpE*), blocking energy homeostasis in both replicating and dormant MTB.<sup>52</sup> Though this new drug has been approved for MDR-TB it remains in Phase III clinical trials for unresolved safety concerns (11.4% of patients taking BDQ died in phase II trials compared to 2.5% taking placebo).<sup>51</sup> Additionally, resistance to BDQ has already been reported. Mutations in *atpE* have been observed as well as mutations in transcriptional regulator Rv0678 which leads to upregulation of efflux pump MmpL.<sup>53-54</sup>

Another investigational drug, delamanid has been approved by the European Union and Japan for the treatment of MDR-TB but has yet to receive approval from the FDA. Delamanid (DLM) is a nitroimidazole which shows efficacy against both actively replicating and latent MTB.<sup>46, 55</sup> DLM acts by interfering with mycolic acid biosynthesis and is activated similarly to pretomanid. Both are prodrugs which are activated by deazaflavin-dependent nitroreductase enzyme (Ddn).<sup>55</sup> Resistance to DLM has been reported to be due to mutations in both the Ddn and a coenzyme F420 gene (*fdg1*).<sup>53</sup>

There are also several other investigational drugs in the TB pipeline which are in early clinical development (see **Figure I-3**). Sutezoid, an oxazolidinone, which targets the ribosome, is in Phase II clinical trials. Other new treatments include Q203, an imidazopyridine and PBTZ169, a piperazine-containing benzothiazinone, both of which are in Phase I clinical trials.<sup>9, 56</sup> Many of the occupants within the pipeline are not novel drugs but new combinations of currently utilized drugs (see **Figure I-3**). Though some

progress has been made in identifying new treatments for drug susceptible and resistant TB, there remains an unmet need for novel treatments.



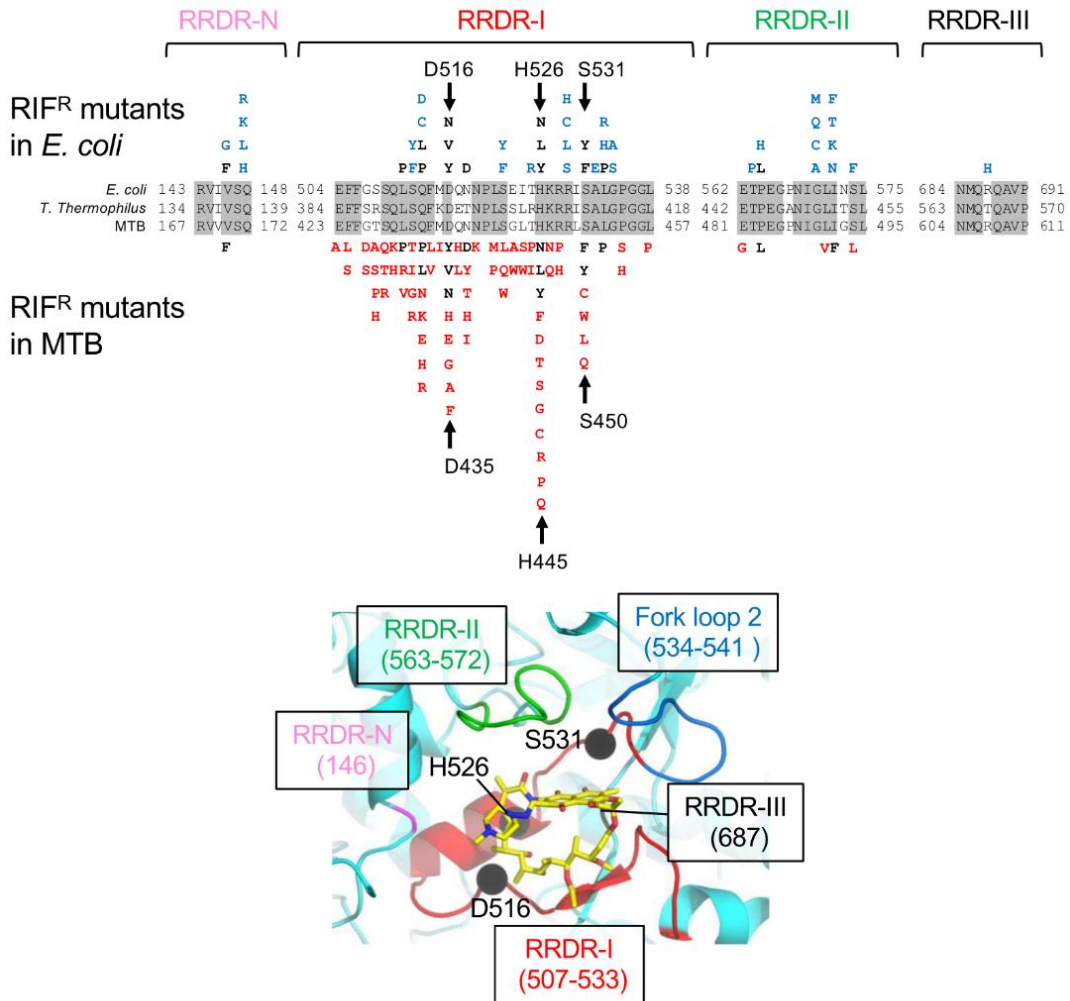
GLP tox=good laboratory practice toxicology studies. DS-TB=drug-sensitive tuberculosis. OBR=optimised background regimen. MDR-TB=multidrug-resistant tuberculosis.

**Figure I-3: Pipeline of drugs for the treatment of tuberculosis.**<sup>56</sup>

### Rifampin Resistance

Rifampin (RMP) has been the cornerstone of all TB treatment regimens since its introduction in the 1960's. Rifampin is a semisynthetic derivative of rifamycin SV which has been modified with a 3-(4-methyl-1-piperazinyl)-iminoethyl modification to improve the pharmacokinetic profile.<sup>35</sup> The core rifamycin is composed of a naphthoquinone which has a highly functionalized, poly-hydroxylated ansamycin ring. RMP and rifamycins are highly selective for bacterial DNA-dependent RNA polymerase (RNAP). Bacteria have a single DNA-dependent RNA polymerase (RNAP) responsible for the production of all

RNA within the cell. RNAP core enzyme is composed of five subunits, two  $\alpha$ -subunits (RpoA),  $\beta$ -subunit (RpoB),  $\beta'$ -subunit (RpoC), and the  $\omega$ - subunit (RpoZ).<sup>57</sup> Rifampin targets a cleft in the  $\beta$ -subunit (RpoB) roughly 20Å from the catalytic center.<sup>58</sup> RMP functions by binding to free and initiating RNAP, blocking elongation of nascent RNA 2-3nt from the active center.<sup>58</sup> RMP has an affinity 10,000-fold weaker for the eukaryotic counterpart, making it a selective antibacterial treatment. The affinity of RMP for bacterial RNAP is in the range of  $10^{-9}$ - $10^{-8}$  M.<sup>59</sup> The surface of the binding cleft in RpoB consists of several loops within the 81-residue rifamycin resistance-determining region (RRDR) and Fork Loop 2 (see **Figure I-4**). The RRDR can be divided into 4 regions: the N-terminal cluster (residue 146, *E. coli* numbering), cluster I (residues 507-533), cluster II (residues 563-572), and cluster III (residue 678).<sup>59</sup>



**Figure I-4: Sequence alignment spanning the RRDRs of the *E. coli*, *T. thermophilus* and MTB  $\beta$ -subunits of RNAP.** RRDRs are indicated above the amino acid sequences. Amino acids that are identical among the three species are shown as gray background. Mutations that confer RMP<sup>R</sup> in *E. coli* and MTB are indicated.<sup>60-61</sup> Three major RMP<sup>R</sup> mutation sites are labeled. Mutations unique for *E. coli* RNAP are shown in blue, mutations unique for MTB RNAP are shown in red, and mutations found in both RNAPs are shown in black (upper panel). The structure of the RMP binding pocket (lower panel).

A vast majority of RMP resistant mutations occur within the RRDR (95%) on the RpoB (see **Figure I-4**). Roughly 90 non-synonymous mutations at 33 codons have been observed in clinical isolates of MTB with RMP<sup>R</sup>.<sup>59, 62-63</sup> However, amino acid substitutions at three codon locations  $\beta$ D435V (MTB numbering),  $\beta$ H445Y, and  $\beta$ S450L comprise 85% of RMP<sup>R</sup> mutants enriched in clinical isolates.<sup>59, 64</sup> Within these three genotypes the overwhelming majority of isolates carry the  $\beta$ S450L mutation (40-70%).<sup>62-63, 65-66</sup> Several biochemical studies have been carried out to elucidate the effects of RMP<sup>R</sup> with *E. coli* RNAP. Laboratory evolved strains of *E. coli* RNAP carrying RMP<sup>R</sup> mutations have been produced; however, the profile of RMP<sup>R</sup> mutations differs between *E. coli* and MTB (see **Figure I-4**).<sup>59</sup> For example, the  $\beta$ S450L mutation is not observed in *E. coli* isolates conferring RMP<sup>R</sup><sup>59</sup>; however, in *E. coli* a serine to phenylalanine or tyrosine substitution has been observed. One explanation for this observation is for *E. coli* to evolve Ser to Leu it would require two nucleotide substitutions (TCT to TTG/TTA) while in MTB only a single point mutation is required (TCG to TTG).<sup>67</sup> Of all the 4 nitrogenous bases cytosine is the most predisposed to chemical modification. This occurs via deamination to uracil which can lead to a C:G to T:A transition.<sup>68</sup> These types of mutations are highly representative of single point mutations.<sup>68</sup> Identification of these specific codon changes is the current method for clinicians to diagnose RMP<sup>R</sup> in clinical isolates via PCR.<sup>69</sup>

Mutations found in RRDR and their effects on RNAP function have been extensively studied in *E. coli*. Many studies used phenotypic screening methods to enrich mutants which perturb enzyme function, such as ability of the enzyme to initiate, elongate, and terminate transcription.<sup>60, 70-71</sup> Additionally, RMP<sup>R</sup> mutations cause pleiotropic effects on cellular physiology and metabolism.<sup>72</sup> Several secondary metabolism genes are

differentially regulated in RMP<sup>R</sup> mutant MTB.<sup>73</sup> For example, polyketide synthase genes *ppsA*, *ppsE*, and *drrA* which produce phthiocerol dimycocerosate as well as other lipids, which are components of the outer cell wall, are overexpressed.<sup>74-75</sup> Altering the composition of the cell wall has been shown to alter the permeability of the cell wall.<sup>76</sup> Rifampin resistance is a complex phenomenon which has far reaching effects for both the patient and the organism itself and remains a highly studied topic to this day.

#### *Fitness Defects and Compensatory Mutations*

Fitness is determined by an organism's ability to survive and replicate (in the case of MTB) under a specific set of conditions. Mutations which impart antibiotic resistance to the organism can often have deleterious effects on the organism's overall fitness. The most prevalent mutation associated with rifamycin resistance is  $\beta$ S450L in the  $\beta$ -subunit. It has been observed that the fitness defects associated with the  $\beta$ S450L RMP<sup>R</sup> mutation vary dramatically between individual isolates, despite carrying the same resistance mutation.<sup>77-80</sup> This led to the suggestion that there may be other mutations present within RNAP which were responsible for fitness variability. Typically, clinicians and research scientists focused on sequencing the RRDR when studying RMP<sup>R</sup> MTB, so until recently, there was little information of the genetic nature of other regions of RNAP. Full gene sequencing of the subunits of RNAP, most notably the  $\beta'$ -subunit, led to discovery of secondary mutations which were associated with RMP<sup>R</sup> mutations (see **Figure I-5**).<sup>66, 80-</sup>



The genome sequences of 117 multiple-drug resistant (MDR) MTB, strains which exhibit resistance to RMP and isoniazid, and 212 MDR-MTB strains from high-burden countries (Georgia, Uzbekistan, and Kazakhstan) were analyzed for the presence of mutations in RNAP subunits other than *rpoB*. Results showed that 30% of the 225 isolates that carried the  $\beta$ S450L mutation had an additional mutation in the  $\beta'$ -subunit (*rpoC*).<sup>82</sup>

The presence of these compensatory mutations is overwhelmingly, but not exclusively associated with the  $\beta$ S450L RMP<sup>R</sup> mutation. Roughly 9% of secondary mutations were observed in the  $\alpha$ -subunit (*rpoA*), though they all localize in proximity to each other spatially.<sup>82</sup> Similar results were also observed in two other studies, which found that 44% and 54% of isolates from West Cape, South Africa and South Korea, respectively, carried a mutation in the  $\beta'$ -subunit/*rpoC* in addition to the  $\beta$ S450L mutation.<sup>66</sup> In consecutive isolates obtained from the three separate patients who had progressed from MDR- to XDR-MTB over a 6-month period, it was found that all isolates from these three patients contained the  $\beta$ S450L mutation with another mutation,  $\beta$ 'V483G.<sup>66</sup> This suggests that these mutations are participating in the progression drug resistance in MTB.

The  $\beta$ 'V483G mutation is the most prevalent mutation associated with the  $\beta$ S450L mutant and has been observed in all of the studies mentioned above, suggesting it is arising by convergent evolution and is physiologically relevant to the organism (see **Figure I-5**).<sup>66, 82</sup> Allelic exchange was used to replace the *rpoB* and *rpoC* genes in *M. smegmatis* (MSG) with their respective counterparts from MTB containing the  $\beta$ S450L and  $\beta$ 'V483G or  $\beta$ 'F452L. The study revealed that the fitness defect associated with the  $\beta$ S450L mutant was fully ameliorated with the  $\beta$ 'V483G mutation and partially



compensated by the  $\beta'$ F452L mutation and this effect was most dramatic under carbon-limiting conditions.<sup>66</sup> The effect observed under carbon-limiting conditions coupled with the localization of the mutations at the double-psi  $\beta$ -barrel, where ppGpp binds suggested that these mutations may be involved in the stringent response (see **Figure I-5**).

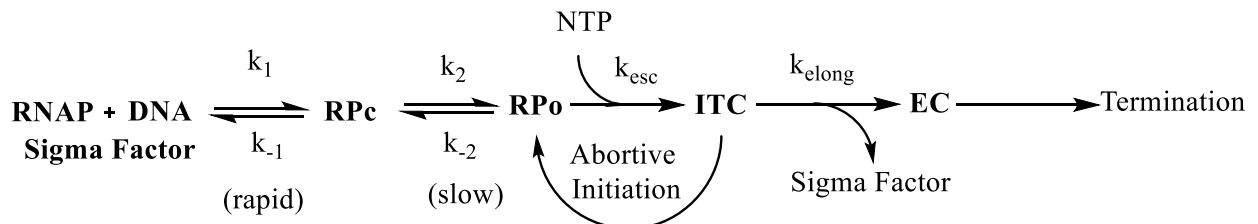
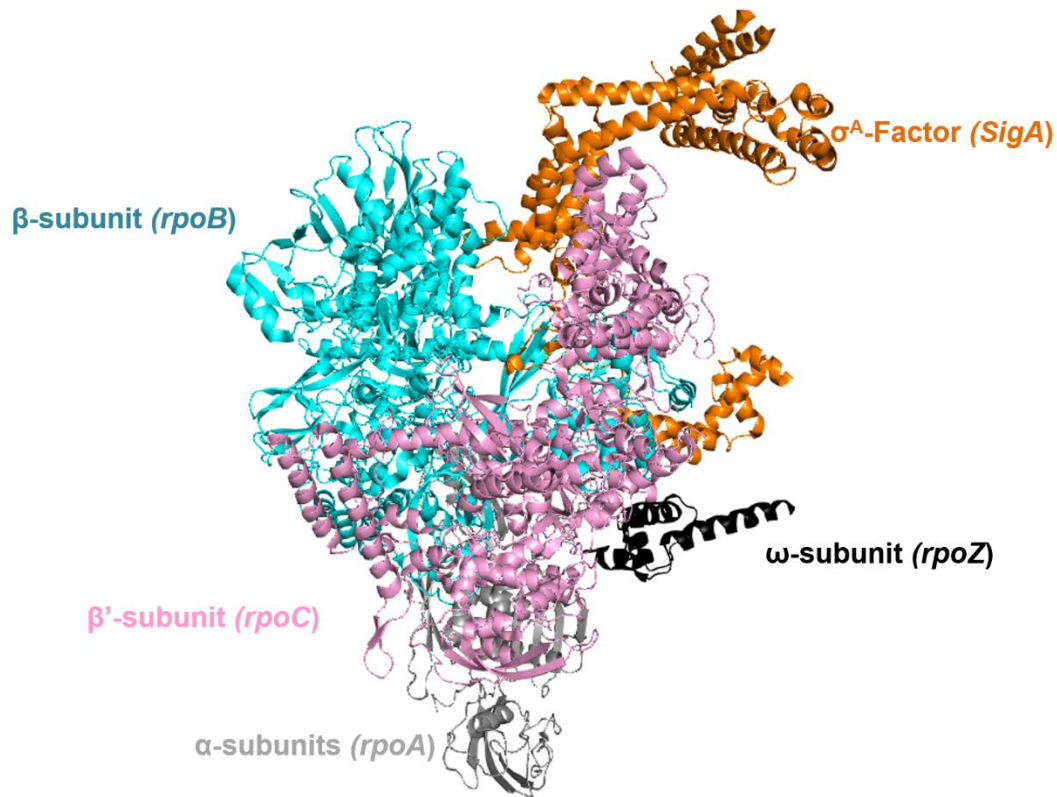
### *RNA polymerase Structure and Function*

As mentioned above, the bacterial RNAP core enzyme is composed of five subunits, two  $\alpha$ , one each of  $\beta$ ,  $\beta'$ , and  $\omega$ .<sup>57</sup> In MTB there are 13 sigma factors (A-M) which, when associated with core enzyme, form the holoenzyme (see **Figure I-6** upper panel).<sup>83</sup> RNAP has been likened to a crab claw, which is composed of a clamp region provided by the  $\beta'$ -subunit with the  $\beta$ -subunit forming the pincer. The clamp is dynamic region of the protein which can adopt both “open” and “closed” states.<sup>57, 84</sup> RNAP is a highly complex and dynamic enzyme which requires several complex mechanistic steps to occur for the production of RNA (see **Figure I-6** lower panel).

With a cognate sigma factor, RNAP binds promoter DNA, which enters through the main channel of RNAP formed by both the  $\beta$ - and  $\beta'$ -subunits (Figure I-5). Core promoter cis-elements consist of a -10, extended -10, and -35 recognition elements (numbers denoting location from the transcription start site or TSS), which are recognized by sigma factor region 2.4 ( $\sigma$ 2.4), sigma factor region 2.5 ( $\sigma$ 2.5), and sigma factor region 4.2 ( $\sigma$ 4.2) respectively.<sup>57</sup> Once RNAP is positioned over the promoter region several isomerization steps occur to go from the closed-promoter complex (RP<sub>c</sub>) to the open-promoter complex (RP<sub>o</sub>), during which promoter DNA is unwound to form the transcription bubble (see **Figure I-6** lower panel). The transcription bubble extends from +1 to -11 relative to the TSS. The RP<sub>o</sub> stability varies from seconds to minutes depending on several factors

including sequence and the presence of secondary transcription factors. At this point RNAP is ready to receive nucleotide triphosphates (NTPs), which enter through the secondary channel.<sup>85-87</sup> As RNAP begins to incorporate NTP into the elongating RNA, the initial transcription complex can go through several abortive initiation events where short RNA's are released.<sup>88</sup> During this time the sigma factor is retained. Abortive initiation is thought to be caused by strong interactions between RNAP and promoter sequences.<sup>88</sup> DNA "scrunching" occurs during abortive initiation, during this process RNAP pulls and unwinds DNA into the main channel inducing stress which is thought to overcome strong interactions with promoter DNA.<sup>89</sup>

After RNAP produces a transcript of roughly 12-15 nucleotides, a structural change occurs which are thought to result in the loss of the sigma factor and leads to stabilization of the RNAP-DNA-RNA ternary complex and productive RNA synthesis, this is referred to the elongation complex (EC).<sup>90</sup> (The loss of the sigma factor upon RNAP transition to the elongation complex remains contentious, FRET experiments have shown that the sigma factor can remain bound to the EC.<sup>91</sup>) EC's are highly processive and stable complexes which are resistant to challenge by high salt, high temperature conditions, as well as, challenge by DNA mimetics such as heparin.<sup>92-93</sup> RNAP can elongate for several thousand bases before recognition of a termination signal. Bacteria contain both rho-dependent and independent termination mechanisms.<sup>94</sup>



**Figure I-6: Structure of RNAP and kinetics of transcription.** MTB RNAP holoenzyme (PDB: 5HUA; upper panel). Kinetic steps required for production of mature RNA (lower panel).

Structurally, RNAPs have several notable subdomains and structures which aid in nucleotide addition. The active center is located at the base of the main channel and is where the catalytic  $Mg^{2+}$  is chelated to three highly conserved aspartic acid residues which are part of a highly conserved DGDbD motif (b = bulky residue) on the double-psi beta-barrel (see **Figure I-5**).<sup>95</sup> The trigger loop (TL) which lies at the center of the main channel can adopt multiple conformations which oscillate as nucleotides are incorporated into the elongating RNA.<sup>96-97</sup> The TL locks nucleotides in the active site, allowing for the

formation of the phosphodiester bond between free NTP and the 3'-OH of the elongating RNA.<sup>97</sup> After incorporation the trigger loop retracts allowing for pyrophosphate to leave through the secondary channel. Another motif which plays a significant role in this process is the bridge helix (BH), which bifurcates the main channel.<sup>98-99</sup> The BH has been observed in both a bent and straight conformation in elongating RNAP's and is thought to work in concert with the trigger loop during nucleotide incorporation.<sup>100</sup> Lastly the fork domain, which is composed of fork loop 1 and fork loop 2, plays a critical role in nucleotide incorporation and processive transcription.<sup>101</sup> The role of fork region has been debated; however, it is clear that it plays a critical role in initiation and elongation.<sup>102</sup> Fork loop 2 is of particular interest as it adopts several conformations and has been proposed to have roles from unwinding DNA to stimulating incorporation of NTPs via fork loop 1.<sup>101-102</sup>

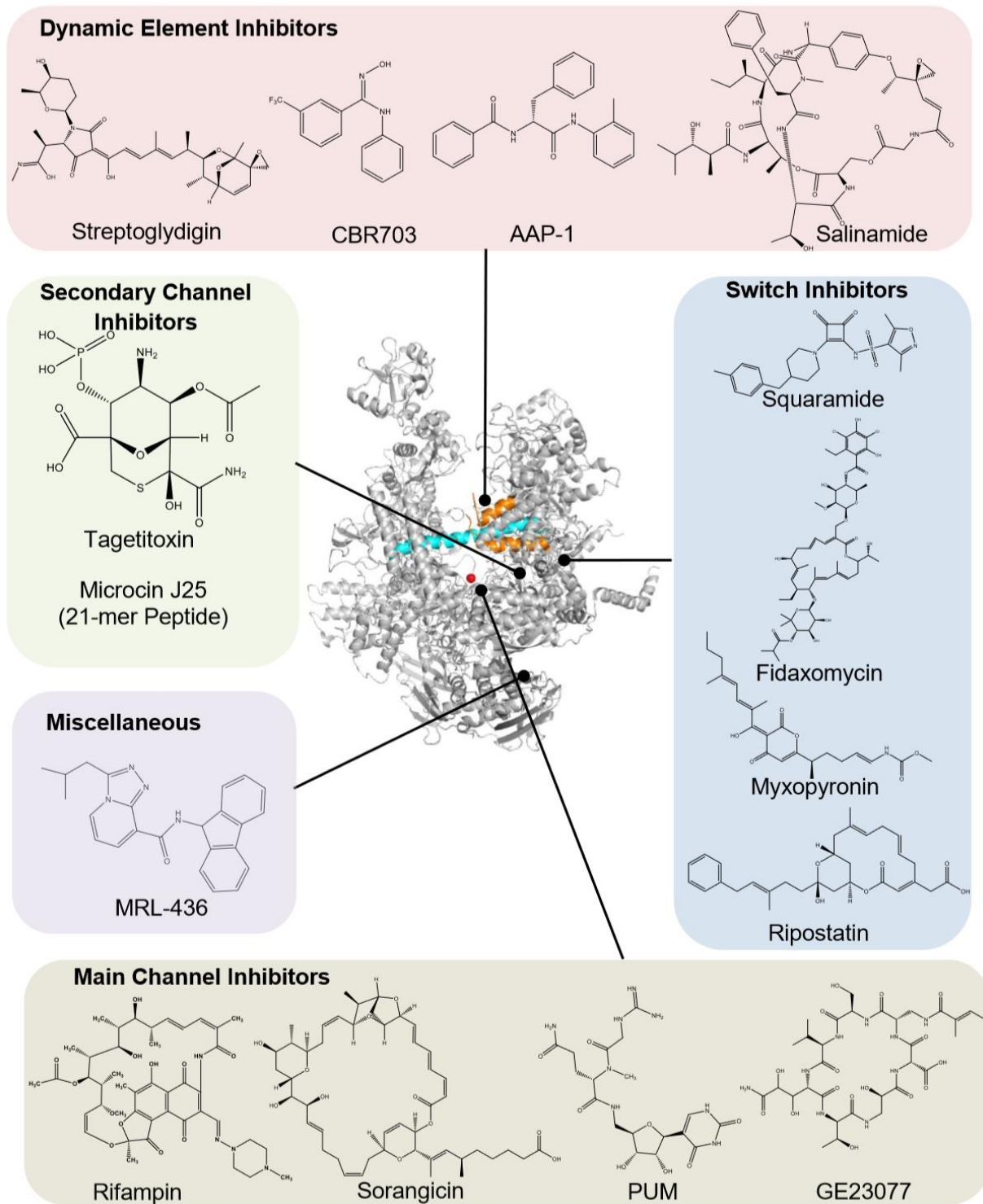
#### *Inhibitors of Bacterial RNA polymerase*

RNAP is a proven effective therapeutic target, as demonstrated by the success of rifampin. There are several reasons for this which include: 1) there is a single RNAP employed by bacteria, which if successfully targeted prevents all RNA related function within the cell, 2) RNAP's are highly conserved among bacteria, which allows for potential broad-spectrum antibiotic activity, and 3) bacterial RNAP's share little sequence similarity to eukaryotic RNAPs allowing for selectivity. Currently there is only one other RNAP inhibitor on the market in addition to rifampin, fidaxomicin (also known as lipiarmycin), which is limited to treatment of gastrointestinal infections caused by gram positive pathogens (primarily *Clostridium difficile*) due to poor systemic bioavailability. Although there are few marketed RNAP inhibitors there are a significant number of bacterial RNAP

inhibitors which have wide ranging mechanisms of action and spatially distinct binding sites within RNAP (see **Figure I-7**).

The switch region of RNAP which aids in clamp closure around DNA is the site of binding for several known RNAP inhibitors (see **Figure I-7**).<sup>84, 103</sup> The function of the switch domain gives a clue to the mechanism of action of these inhibitors. Upon binding of RNAP to DNA it is in the “open” conformation, structural changes within the switch domain induce closure of the clamp over the DNA leading to a “closed” conformation.<sup>103</sup> Fidaxomicin is a natural product produced by *Dactylosporangium aurantiacum* which binds within the switch domain preventing the closure of the clamp domain around DNA.<sup>84</sup> *In vitro* characterization of Fidaxomicin has shown that it has a modest effect on Gram negative RNAP (7  $\mu$ M in *E. coli* and MTB).<sup>104</sup> However, recent work has shown that the transcription factor RNA polymerase binding protein (RbpA) is required for high potency inhibition with Gram negative bacterial RNAP (0.21  $\mu$ M).<sup>104</sup>

There are a set of switch domain inhibitors which act oppositely to Fidaxomicin, stabilizing the closed conformation of the clamp, preventing double stranded DNA from entering the main channel. Myxopyronin, coralopyronin, and ripostatin bind a hydrophobic cleft near switch 1 and switch 2, locking the clamp in the closed conformation.<sup>103</sup> Though ripostatin is structurally dissimilar to both myxopyronin and coralopyronin they share the same binding cleft.<sup>103</sup> Another set of molecules which bind a similar region of RNAP are the squaramides. The squaramides were identified in a coupled transcription-translation assay.<sup>105</sup> These compounds act by shifting switch 2 into the main channel, disrupting the placement of melted DNA.<sup>106</sup>



**Figure I-7: Inhibitors of RNA polymerase and functional region of RNAP which they target.**

The main channel of RNAP is host to many inhibitors as well (see **Figure I-7**). Sorangicin, a macrocyclic compound isolated from *Sorangium cellulosum* occupies the same binding site as rifampin.<sup>107-108</sup> Adjacent to the rifampin site, in the direction of the active site is where the cyclic peptide natural product GE23077 binds.<sup>109</sup> This macrocyclic heptapeptide occupies the *i* and *i* + 1 sites. Its large size has limited its efficacy due to issues with permeability, but conjugates with rifampin showed efficacy against both wild-type and RMP<sup>R</sup> mutant RNAPs in vitro, which suggests if permeability issues could be resolved this could be a promising lead for the treatment of RMP<sup>R</sup> MTB.<sup>109</sup> Recently, the nucleoside analog inhibitor, pseudouridimycin (PUM), was identified from a microbial-extract screen. This inhibitor occupies the active site and competes with UTP.<sup>110</sup>

Within the main channel there are several dynamic motifs, including the bridge helix and trigger loop, and these also are binding sites for several RNAP inhibitors (see **Figure I-7**). Streptolydigin is a natural product which targets elongating RNAPs. It binds adjacent to the bridge helix and trigger loop preventing conformation changes of both required for nucleotide incorporation.<sup>111</sup> Salinamide A and E (Sal) are bicyclic polypeptides which were isolated from *Streptomyces spp.* and target RNAPs from both Gram negative and positive pathogens. Sal binds in a region between the secondary channel and the bridge helix and is thought to disrupt the function of the bridge helix, preventing nucleotide addition.<sup>112</sup> The CBR series of small molecules were identified in an *in vitro* *E. coli* transcription screen. These small molecules stabilize the elongation complex slowing down translocation by allosterically modulating both the trigger loop and bridge helix.<sup>113</sup> N $\alpha$ -aroyl-N-phenylalaninamides (AAP's), identified in a separate screen, bind similarly to the CBR analogs.<sup>114</sup>

There are several inhibitors which target the secondary channel as well as other sites on RNAP (see **Figure I-7**). Tagetitoxin, a natural product isolated from *Pseudomonas syringae* pv. *Tagetis*, shows activity for both prokaryotic and eukaryotic RNAPs.<sup>115</sup> A structure of tagetitoxin with *T. thermophilus* shows that it binds the secondary channel, however it has been proposed to also interact with the trigger loop.<sup>116</sup> Microcin J25 is a 21-mer peptide which prevents NTPs from entering the active site, shows activity against *E. coli* RNAP.<sup>117</sup> MRL-436 is a small molecule identified from a mass spectrometry-based screen.<sup>118</sup> MRL-436 has been shown to bind to the same site as ppGpp, a small molecule involved in the stringent response, adjacent to the  $\omega$ -subunit.<sup>118</sup> The exact mechanism of this molecule is still unknown; however, because of its proximity to the ppGpp site suggests it potentially destabilizes the open-promoter complex.<sup>118</sup>

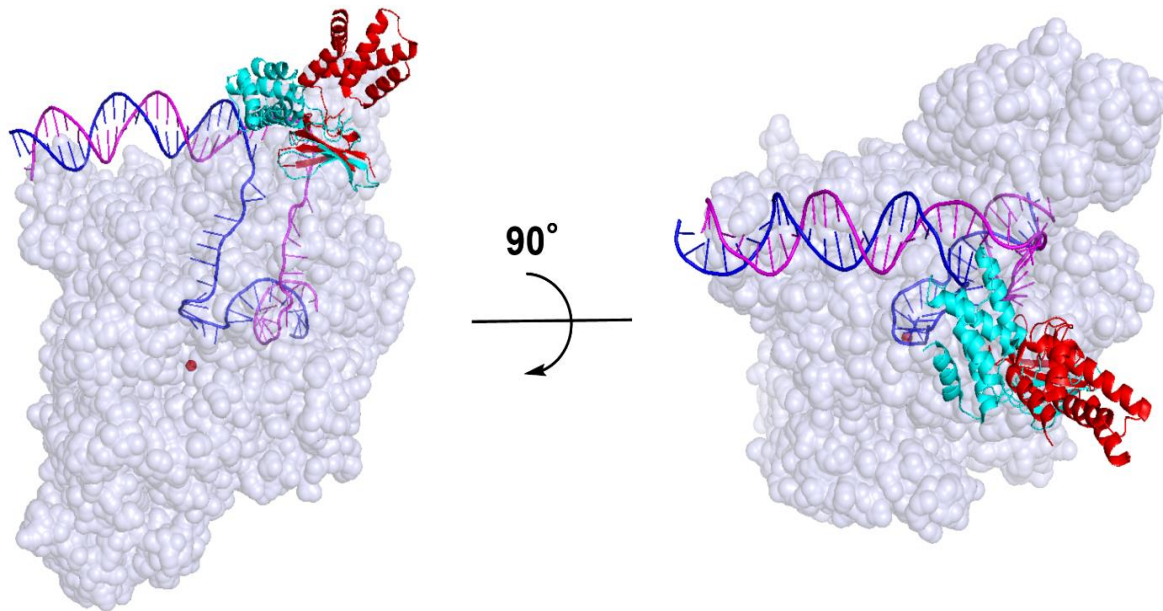
All of the inhibitors mentioned up to this point have targeted core RNAP (see **Figure I-7**). Several inhibitors have been identified which target the interaction between core RNAP and the sigma factor. DSHS00507 was identified in *in silico* screening was shown to prevent sigma factor association with RNAP by binding to the  $\beta'$ -subunit CH region.<sup>119</sup> The SB small molecules series of small molecules was identified in a ELISA based assay using *E. coli* RNAP were also shown to disrupt holoenzyme formation.<sup>120</sup> Lastly the GLK series of small molecules which bind to the same motif as the SB series were identified via a structure-based design using amino acid interactions between RNAP and the sigma factor to generate a pharmacophore model.<sup>119, 121</sup> RNAP has been extensively targeted for drug development; however, targeting of trans-acting factors which contribute to the function of RNAP has not be extensively attempted. The only



transcription factor which has been successfully targeted is rho, a termination factor. Bicyclomycin, a natural product derived from *Streptomyces sapporonensis*, prevents Rho-dependent termination and ATP hydrolysis.<sup>122</sup> The development of therapeutic agents which target transcription factors represents a gap in the efforts to inhibit transcription.

### *Mycobacterial Transcription Regulator CarD*

CarD is a global transcriptional regulator found in certain groups of both Gram positive- and Gram-negative bacteria.<sup>85, 123-125</sup> After RNAP binds to a promoter, forming the closed complex (RPc), the transcription bubble/open promoter complex (RPO) is formed by unwinding the DNA allowing RNA to be transcribed. This step is thought to be the rate limiting step for gene expression. CarD stabilizes the open-promoter complex.<sup>85, 126-127</sup> Stabilization of this complex is essential for MTB viability; attempts to remove CarD from the MTB genome were lethal.<sup>128-130</sup> However, conditional depletion of CarD using a tetracycline-controlled gene expression led to increased sensitivity to reactive oxygen species and RMP, as well as the inability of MTB to replicate and persist in mice.<sup>128</sup> Many of the mutations in CarD which were attempted to be introduced through homologous recombination were not tolerated which suggests that small changes in protein structure have a great impact on CarD function and MTB viability.



**Figure I-8: Interaction of CarD with RNAP.** Structure of *T. aquaticus* (Taq) RNAP (PDB: 4XLR) open-promoter complex with Taq CarD (Cyan) superimposed with MTB β1β2-lobe (hidden) and MTB CarD (red PDB: 4KBM). The structure is rotated out of the plane of the paper to show the contacts of the Taq CarD (Cyan) with the upstream fork junction of the transcription bubble. MTB is in the “out” conformation and Taq CarD in the “in” conformation, making contacts with both RNAP and promoter DNA.

The N-terminal domain of CarD interacts with the β1β2-lobes on the β-subunit of RNAP, which is proximal to Rif binding site (see **Figure I-8**). Crystallographic data suggests that a conformational change induced by CarD binding may modulate RMP affinity.<sup>131</sup> Additionally, weakening of the interaction between CarD and RNAP resulted in increased MTB susceptibility to RMP, consistent with a conformational linkage between the CarD and Rif binding sites.<sup>130</sup> This may suggest that development of therapeutics which inhibit the CarD•RNAP interaction might also improve RMP efficacy.

CarD homologues are present in many bacterial phyla, making it an attractive therapeutic target.<sup>85</sup> Notable Phyla (disease states) which carry CarD proteins include: *Chlamydiae* (chlamydia), *Firmicutes* (*C. diff.* colitis), *Rickettsia* (typhus) and *Spirochaetes*

(Lyme disease) in addition to other clinically relevant pathogens. With this in mind, it is our belief that CarD represents an appealing target for therapeutic discovery for the treatment of tuberculosis as well as other clinically relevant pathogens.

### **Research Objectives**

MTB RNAP polymerase represents an attractive and proven therapeutic target for novel antibiotic discovery. The main goals of this dissertation were three-fold: 1) to obtain highly pure and active recombinant MTB RNAP, 2) to use this purified enzyme to characterize clinically relevant MTB RNAP's both mechanistically and kinetically, to gain a better understanding of rifampin-resistance, and 3) to look for novel approaches to target transcription in MTB via CarD, a transcriptional regulator, for future novel therapeutic potential. Over the course of this project, we have gained insight into MTB RNAP function and have been successful in priming and developing CarD for therapeutic discovery. The primary objectives described in each chapter are as follows:

#### **Chapter 2: To overexpress, purify and characterize MTB RNA polymerase.**

Using multiple expression plasmids, an *in vivo* reconstituted MTB RNAP purification was optimized for the recovery of highly pure and active enzyme. Purified enzyme was used to identify the IC<sub>50</sub> of rifampin for WT and RMP<sup>R</sup> MTB RNAPs. This work was complemented with structural information about the RMP<sup>R</sup> mutant RNAPs using *E. coli* RNAP.

### **Chapter 3: To characterize clinically relevant Rifampin Resistant *M. tuberculosis* RNA polymerases *In vitro***

Using purified MTB RNAP enzymes several *in vitro* RNA transcription assays were utilized to probe initiation, elongation, termination, and 3'-RNA hydrolysis with the wild-type and RMP<sup>R</sup> MTB RNAPs. We found that the less prevalent RMP<sup>R</sup> mutants exhibit significantly poorer termination efficiencies relative to wild-type, an important factor for proper gene expression. We also found that several mechanistic aspects of transcription of the RMP<sup>R</sup> mutant RNAPs are impacted relative to wild-type.

### **Chapter 4: To development, optimize and execute a Fluorescence Polarization assay to Identify Inhibitors of MTB Transcriptional Regulator CarD via HTS**

A fluorescence polarization to monitor the binding of MTB transcriptional regulator CarD to the RNAP•DNA ternary complex was developed and optimized. Using this assay both a 237 and 23,320 compound HTS were conducted. Hits were characterized using biochemical, biophysical, and bacterial techniques to identify novel inhibitors of both the CarD interaction with the RNAP•DNA ternary complex and inhibitors of DNA association with RNAP.

## References

1. Brown, E. D.; Wright, G. D., Antibacterial drug discovery in the resistance era. *Nature* **2016**, *529* (7586), 336-43.
2. Zaman, S. B.; Hussain, M. A.; Nye, R.; Mehta, V.; Mamun, K. T.; Hossain, N., A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **2017**, *9* (6), e1403.
3. Munita, J. M.; Arias, C. A., Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **2016**, *4* (2).
4. Prevention, C. f. D. C. a.
5. Hay, M.; Thomas, D. W.; Craighead, J. L.; Economides, C.; Rosenthal, J., Clinical development success rates for investigational drugs. *Nat Biotechnol* **2014**, *32* (1), 40-51.
6. Organization, W. H. *Global tuberculosis Report 2016*; 2017.
7. Cadena, A. M.; Fortune, S. M.; Flynn, J. L., Heterogeneity in tuberculosis. *Nat Rev Immunol* **2017**, *17* (11), 691-702.
8. Philips, J. A.; Ernst, J. D., Tuberculosis pathogenesis and immunity. *Annu Rev Pathol* **2012**, *7*, 353-84.
9. Zumla, A.; Chakaya, J.; Centis, R.; D'Ambrosio, L.; Mwaba, P.; Bates, M.; Kapata, N.; Nyirenda, T.; Chanda, D.; Mfinanga, S.; Hoelscher, M.; Maeurer, M.; Migliori, G. B., Tuberculosis treatment and management--an update on treatment regimens, trials, new drugs, and adjunct therapies. *Lancet Respir Med* **2015**, *3* (3), 220-34.
10. Zimmerli, S.; Edwards, S.; Ernst, J. D., Selective receptor blockade during phagocytosis does not alter the survival and growth of Mycobacterium tuberculosis in human macrophages. *Am J Respir Cell Mol Biol* **1996**, *15* (6), 760-70.

11. Kang, P. B.; Azad, A. K.; Torrelles, J. B.; Kaufman, T. M.; Beharka, A.; Tibesar, E.; DesJardin, L. E.; Schlesinger, L. S., The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis. *J Exp Med* **2005**, *202* (7), 987-99.
12. Armstrong, J. A.; Hart, P. D., Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. *J Exp Med* **1975**, *142* (1), 1-16.
13. Philips, J. A., Mycobacterial manipulation of vacuolar sorting. *Cell Microbiol* **2008**, *10* (12), 2408-15.
14. Vergne, I.; Chua, J.; Lee, H. H.; Lucas, M.; Belisle, J.; Deretic, V., Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **2005**, *102* (11), 4033-8.
15. Fratti, R. A.; Backer, J. M.; Gruenberg, J.; Corvera, S.; Deretic, V., Role of phosphatidylinositol 3-kinase and Rab5 effectors in phagosomal biogenesis and mycobacterial phagosome maturation arrest. *J Cell Biol* **2001**, *154* (3), 631-44.
16. van der Wel, N.; Hava, D.; Houben, D.; Fluitsma, D.; van Zon, M.; Pierson, J.; Brenner, M.; Peters, P. J., M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* **2007**, *129* (7), 1287-98.
17. Smith, J.; Manoranjan, J.; Pan, M.; Bohsali, A.; Xu, J.; Liu, J.; McDonald, K. L.; Szyk, A.; LaRonde-LeBlanc, N.; Gao, L. Y., Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in Mycobacterium marinum escape from the vacuole. *Infect Immun* **2008**, *76* (12), 5478-87.
18. Delgado, M. A.; Elmaoued, R. A.; Davis, A. S.; Kyei, G.; Deretic, V., Toll-like receptors control autophagy. *EMBO J* **2008**, *27* (7), 1110-21.
19. MacMicking, J. D.; Taylor, G. A.; McKinney, J. D., Immune control of tuberculosis by IFN-gamma-inducible LRG-47. *Science* **2003**, *302* (5645), 654-9.
20. Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. A.; Wang, S., Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **2004**, *332* (2), 261-73.

21. Tsai, M. C.; Chakravarty, S.; Zhu, G.; Xu, J.; Tanaka, K.; Koch, C.; Tufariello, J.; Flynn, J.; Chan, J., Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. *Cell Microbiol* **2006**, *8* (2), 218-32.
22. Flynn, J. L.; Chan, J.; Lin, P. L., Macrophages and control of granulomatous inflammation in tuberculosis. *Mucosal Immunol* **2011**, *4* (3), 271-8.
23. Gengenbacher, M.; Kaufmann, S. H., Mycobacterium tuberculosis: success through dormancy. *FEMS Microbiol Rev* **2012**, *36* (3), 514-32.
24. Primm, T. P.; Andersen, S. J.; Mizrahi, V.; Avarbock, D.; Rubin, H.; Barry, C. E., 3rd, The stringent response of Mycobacterium tuberculosis is required for long-term survival. *J Bacteriol* **2000**, *182* (17), 4889-98.
25. Haurlyliuk, V.; Atkinson, G. C.; Murakami, K. S.; Tenson, T.; Gerdes, K., Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **2015**, *13* (5), 298-309.
26. Daniel, T. M., The history of tuberculosis. *Respir Med* **2006**, *100* (11), 1862-70.
27. Daniel, T. M., Robert Koch and the pathogenesis of tuberculosis. *Int J Tuberc Lung Dis* **2005**, *9* (11), 1181-2.
28. Daniel, T. M., Leon Charles Albert Calmette and BCG vaccine. *Int J Tuberc Lung Dis* **2005**, *9* (9), 944-5.
29. Comstock, G. W., The International Tuberculosis Campaign: a pioneering venture in mass vaccination and research. *Clin Infect Dis* **1994**, *19* (3), 528-40.
30. Cohn, D. L., The effect of BCG vaccination on tuberculin skin testing. Does it matter? *Am J Respir Crit Care Med* **2001**, *164* (6), 915-6.
31. Schatz, A.; Bugie, E.; Waksman, S. A., Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. *Clin Orthop Relat Res* **2005**, (437), 3-6.

32. Sensi, P., History of the development of rifampin. *Rev Infect Dis* **1983**, 5 Suppl 3, S402-6.
33. Chakraborty, S.; Rhee, K. Y., Tuberculosis Drug Development: History and Evolution of the Mechanism-Based Paradigm. *Cold Spring Harb Perspect Med* **2015**, 5 (8), a021147.
34. Horsburgh, C. R., Jr.; Barry, C. E., 3rd; Lange, C., Treatment of Tuberculosis. *N Engl J Med* **2015**, 373 (22), 2149-60.
35. Aristoff, P. A.; Garcia, G. A.; Kirchhoff, P. D.; Showalter, H. D., Rifamycins--obstacles and opportunities. *Tuberculosis (Edinb)* **2010**, 90 (2), 94-118.
36. Mitchison, D. A., Role of individual drugs in the chemotherapy of tuberculosis. *Int J Tuberc Lung Dis* **2000**, 4 (9), 796-806.
37. Fernandes, G.; Salgado, H. R. N.; Santos, J. L. D., Isoniazid: A Review of Characteristics, Properties and Analytical Methods. *Crit Rev Anal Chem* **2017**, 47 (4), 298-308.
38. Unissa, A. N.; Subbian, S.; Hanna, L. E.; Selvakumar, N., Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis. *Infect Genet Evol* **2016**, 45, 474-492.
39. David E. Minnikin, O. Y.-C. L., Houdini H.T. Wu, Vijayashankar Nataraj, Helen D. Donoghue, Malin Ridell, Motoko Watanabe, Luke Alderwick, Apoorva Bhatt and Gurdyal S. Besra, Pathophysiological Implications of Cell Envelope Structure in Mycobacterium tuberculosis and Related Taxa, Tuberculosis. *IntechOpen* **2015**.
40. Zhang, Y.; Mitchison, D., The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis* **2003**, 7 (1), 6-21.
41. Sheen, P.; Lozano, K.; Gilman, R. H.; Valencia, H. J.; Loli, S.; Fuentes, P.; Grandjean, L.; Zimic, M., pncA gene expression and prediction factors on pyrazinamide resistance in Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **2013**, 93 (5), 515-22.



42. Goude, R.; Amin, A. G.; Chatterjee, D.; Parish, T., The arabinosyltransferase EmbC is inhibited by ethambutol in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **2009**, *53* (10), 4138-46.
43. Donald, P. R.; Maher, D.; Maritz, J. S.; Qazi, S., Ethambutol dosage for the treatment of children: literature review and recommendations. *Int J Tuberc Lung Dis* **2006**, *10* (12), 1318-30.
44. Zhao, L. L.; Sun, Q.; Liu, H. C.; Wu, X. C.; Xiao, T. Y.; Zhao, X. Q.; Li, G. L.; Jiang, Y.; Zeng, C. Y.; Wan, K. L., Analysis of embCAB mutations associated with ethambutol resistance in multidrug-resistant mycobacterium tuberculosis isolates from China. *Antimicrob Agents Chemother* **2015**, *59* (4), 2045-50.
45. Marks, S. M.; Flood, J.; Seaworth, B.; Hirsch-Moverman, Y.; Armstrong, L.; Mase, S.; Salcedo, K.; Oh, P.; Graviss, E. A.; Colson, P. W.; Armitige, L.; Revuelta, M.; Sheeran, K.; Consortium, T. B. E. S., Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005-2007. *Emerg Infect Dis* **2014**, *20* (5), 812-21.
46. Falzon, D.; Schunemann, H. J.; Harausz, E.; Gonzalez-Angulo, L.; Lienhardt, C.; Jaramillo, E.; Weyer, K., World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *Eur Respir J* **2017**, *49* (3).
47. Kang, B. H.; Jo, K. W.; Shim, T. S., Current Status of Fluoroquinolone Use for Treatment of Tuberculosis in a Tertiary Care Hospital in Korea. *Tuberc Respir Dis (Seoul)* **2017**, *80* (2), 143-152.
48. Redgrave, L. S.; Sutton, S. B.; Webber, M. A.; Piddock, L. J., Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol* **2014**, *22* (8), 438-45.
49. Reeves, A. Z.; Campbell, P. J.; Sultana, R.; Malik, S.; Murray, M.; Plikaytis, B. B.; Shinnick, T. M.; Posey, J. E., Aminoglycoside cross-resistance in *Mycobacterium tuberculosis* due to mutations in the 5' untranslated region of whiB7. *Antimicrob Agents Chemother* **2013**, *57* (4), 1857-65.
50. Zumla, A.; Nahid, P.; Cole, S. T., Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov* **2013**, *12* (5), 388-404.

51. Cox, E.; Laessig, K., FDA approval of bedaquiline--the benefit-risk balance for drug-resistant tuberculosis. *N Engl J Med* **2014**, *371* (8), 689-91.
52. Hards, K.; Robson, J. R.; Berney, M.; Shaw, L.; Bald, D.; Koul, A.; Andries, K.; Cook, G. M., Bactericidal mode of action of bedaquiline. *J Antimicrob Chemother* **2015**, *70* (7), 2028-37.
53. Bloemberg, G. V.; Keller, P. M.; Stucki, D.; Trauner, A.; Borrell, S.; Latshang, T.; Coscolla, M.; Rothe, T.; Homke, R.; Ritter, C.; Feldmann, J.; Schulthess, B.; Gagneux, S.; Bottger, E. C., Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. *N Engl J Med* **2015**, *373* (20), 1986-8.
54. Xu, J.; Wang, B.; Hu, M.; Huo, F.; Guo, S.; Jing, W.; Nuernberger, E.; Lu, Y., Primary clofazimine and bedaquiline resistance among isolates from patients with multidrug-resistant tuberculosis. *Antimicrob Agents Chemother* **2017**.
55. Xavier, A. S.; Lakshmanan, M., Delamanid: A new armor in combating drug-resistant tuberculosis. *J Pharmacol Pharmacother* **2014**, *5* (3), 222-4.
56. Lohrasbi, V.; Talebi, M.; Bialvaei, A. Z.; Fattorini, L.; Drancourt, M.; Heidary, M.; Darban-Sarokhalil, D., Trends in the discovery of new drugs for Mycobacterium tuberculosis therapy with a glance at resistance. *Tuberculosis (Edinb)* **2018**, *109*, 17-27.
57. Murakami, K. S., X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem* **2013**, *288* (13), 9126-34.
58. Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A., Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **2001**, *104* (6), 901-12.
59. Molodtsov, V.; Scharf, N. T.; Stefan, M. A.; Garcia, G. A.; Murakami, K. S., Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis. *Mol Microbiol* **2017**, *103* (6), 1034-1045.
60. Zhou, Y. N.; Lubkowska, L.; Hui, M.; Court, C.; Chen, S.; Court, D. L.; Strathern, J.; Jin, D. J.; Kashlev, M., Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in Escherichia coli. *J Biol Chem* **2013**, *288* (4), 2700-10.

61. Sandgren, A.; Strong, M.; Muthukrishnan, P.; Weiner, B. K.; Church, G. M.; Murray, M. B., Tuberculosis drug resistance mutation database. *PLoS Med* **2009**, *6* (2), e2.
62. Poudel, A.; Nakajima, C.; Fukushima, Y.; Suzuki, H.; Pandey, B. D.; Maharjan, B.; Suzuki, Y., Molecular characterization of multidrug-resistant *Mycobacterium tuberculosis* isolated in Nepal. *Antimicrob Agents Chemother* **2012**, *56* (6), 2831-6.
63. Tang, K.; Sun, H.; Zhao, Y.; Guo, J.; Zhang, C.; Feng, Q.; He, Y.; Luo, M.; Li, Y.; Sun, Q., Characterization of rifampin-resistant isolates of *Mycobacterium tuberculosis* from Sichuan in China. *Tuberculosis (Edinb)* **2013**, *93* (1), 89-95.
64. Gill, S. K.; Garcia, G. A., Rifamycin inhibition of WT and Rif-resistant *Mycobacterium tuberculosis* and *Escherichia coli* RNA polymerases in vitro. *Tuberculosis (Edinb)* **2011**, *91* (5), 361-9.
65. Rahmo, A.; Hamdar, Z.; Kasaa, I.; Dabboussi, F.; Hamze, M., Genotypic detection of rifampicin-resistant *M. tuberculosis* strains in Syrian and Lebanese patients. *J Infect Public Health* **2012**, *5* (6), 381-7.
66. Song, T.; Park, Y.; Shamputa, I. C.; Seo, S.; Lee, S. Y.; Jeon, H. S.; Choi, H.; Lee, M.; Glynn, R. J.; Barnes, S. W.; Walker, J. R.; Batalov, S.; Yusim, K.; Feng, S.; Tung, C. S.; Theiler, J.; Via, L. E.; Boshoff, H. I.; Murakami, K. S.; Korber, B.; Barry, C. E., 3rd; Cho, S. N., Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the beta' subunit of RNA polymerase. *Mol Microbiol* **2014**, *91* (6), 1106-19.
67. Goldstein, B. P., Resistance to rifampicin: a review. *J Antibiot (Tokyo)* **2014**, *67* (9), 625-30.
68. Lewis, C. A., Jr.; Crayle, J.; Zhou, S.; Swanstrom, R.; Wolfenden, R., Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc Natl Acad Sci U S A* **2016**, *113* (29), 8194-9.
69. Zeka, A. N.; Tasbakan, S.; Cavusoglu, C., Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *J Clin Microbiol* **2011**, *49* (12), 4138-41.

70. Jin, D. J.; Gross, C. A., RpoB8, a rifampicin-resistant termination-proficient RNA polymerase, has an increased Km for purine nucleotides during transcription elongation. *J Biol Chem* **1991**, *266* (22), 14478-85.
71. Jin, D. J.; Walter, W. A.; Gross, C. A., Characterization of the termination phenotypes of rifampicin-resistant mutants. *J Mol Biol* **1988**, *202* (2), 245-53.
72. Hall, A. R.; Iles, J. C.; MacLean, R. C., The fitness cost of rifampicin resistance in *Pseudomonas aeruginosa* depends on demand for RNA polymerase. *Genetics* **2011**, *187* (3), 817-22.
73. Alifano, P.; Palumbo, C.; Pasanisi, D.; Tala, A., Rifampicin-resistance, rpoB polymorphism and RNA polymerase genetic engineering. *J Biotechnol* **2015**, *202*, 60-77.
74. Bisson, G. P.; Mehaffy, C.; Broeckling, C.; Prenni, J.; Rifat, D.; Lun, D. S.; Burgos, M.; Weissman, D.; Karakousis, P. C.; Dobos, K., Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant *Mycobacterium tuberculosis*. *J Bacteriol* **2012**, *194* (23), 6441-52.
75. Cox, J. S.; Chen, B.; McNeil, M.; Jacobs, W. R., Jr., Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **1999**, *402* (6757), 79-83.
76. Camacho, L. R.; Constant, P.; Raynaud, C.; Laneelle, M. A.; Triccas, J. A.; Gicquel, B.; Daffe, M.; Guilhot, C., Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **2001**, *276* (23), 19845-54.
77. Gagneux, S., Fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Clin Microbiol Infect* **2009**, *15 Suppl 1*, 66-8.
78. Billington, O. J.; McHugh, T. D.; Gillespie, S. H., Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **1999**, *43* (8), 1866-9.
79. Mariam, D. H.; Mengistu, Y.; Hoffner, S. E.; Andersson, D. I., Effect of rpoB mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **2004**, *48* (4), 1289-94.

80. Brandis, G.; Hughes, D., Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *J Antimicrob Chemother* **2013**, *68* (11), 2493-7.
81. Meftahi, N.; Namouchi, A.; Mhenni, B.; Brandis, G.; Hughes, D.; Mardassi, H., Evidence for the critical role of a secondary site rpoB mutation in the compensatory evolution and successful transmission of an MDR tuberculosis outbreak strain. *J Antimicrob Chemother* **2016**, *71* (2), 324-32.
82. Comas, I.; Borrell, S.; Roetzer, A.; Rose, G.; Malla, B.; Kato-Maeda, M.; Galagan, J.; Niemann, S.; Gagneux, S., Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* **2011**, *44* (1), 106-10.
83. Sachdeva, P.; Misra, R.; Tyagi, A. K.; Singh, Y., The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J* **2010**, *277* (3), 605-26.
84. Lin, W.; Das, K.; Degen, D.; Mazumder, A.; Duchi, D.; Wang, D.; Ebright, Y. W.; Ebright, R. Y.; Sineva, E.; Gigliotti, M.; Srivastava, A.; Mandal, S.; Jiang, Y.; Liu, Y.; Yin, R.; Zhang, Z.; Eng, E. T.; Thomas, D.; Donadio, S.; Zhang, H.; Zhang, C.; Kapanidis, A. N.; Ebright, R. H., Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Mol Cell* **2018**, *70* (1), 60-71 e15.
85. Bae, B.; Chen, J.; Davis, E.; Leon, K.; Darst, S. A.; Campbell, E. A., CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* **2015**, *4*.
86. Rutherford, S. T.; Villers, C. L.; Lee, J. H.; Ross, W.; Gourse, R. L., Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. *Genes Dev* **2009**, *23* (2), 236-48.
87. Nechaev, S.; Chlenov, M.; Severinov, K., Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. *J Biol Chem* **2000**, *275* (33), 25516-22.
88. Hsu, L. M., Monitoring abortive initiation. *Methods* **2009**, *47* (1), 25-36.

89. Kapanidis, A. N.; Margeat, E.; Ho, S. O.; Kortkhonjia, E.; Weiss, S.; Ebright, R. H., Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **2006**, *314* (5802), 1144-7.
90. Hansen, U. M.; McClure, W. R., Role of the sigma subunit of Escherichia coli RNA polymerase in initiation. II. Release of sigma from ternary complexes. *J Biol Chem* **1980**, *255* (20), 9564-70.
91. Mukhopadhyay, J.; Kapanidis, A. N.; Mekler, V.; Kortkhonjia, E.; Ebright, Y. W.; Ebright, R. H., Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **2001**, *106* (4), 453-63.
92. Wilson, K. S.; Conant, C. R.; von Hippel, P. H., Determinants of the stability of transcription elongation complexes: interactions of the nascent RNA with the DNA template and the RNA polymerase. *J Mol Biol* **1999**, *289* (5), 1179-94.
93. Jin, Z.; Leveque, V.; Ma, H.; Johnson, K. A.; Klumpp, K., Assembly, purification, and pre-steady-state kinetic analysis of active RNA-dependent RNA polymerase elongation complex. *J Biol Chem* **2012**, *287* (13), 10674-83.
94. Peters, J. M.; Vangeloff, A. D.; Landick, R., Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* **2011**, *412* (5), 793-813.
95. Iyer, L. M.; Koonin, E. V.; Aravind, L., Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Structural Biology* **2003**, *3* (1), 1.
96. Zhang, J.; Palangat, M.; Landick, R., Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* **2010**, *17* (1), 99-104.
97. Mishanina, T. V.; Palo, M. Z.; Nayak, D.; Mooney, R. A.; Landick, R., Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci U S A* **2017**, *114* (26), E5103-E5112.
98. Bar-Nahum, G.; Epshtein, V.; Ruckenstein, A. E.; Rafikov, R.; Mustaev, A.; Nudler, E., A ratchet mechanism of transcription elongation and its control. *Cell* **2005**, *120* (2), 183-93.

99. Zhang, N.; Schafer, J.; Sharma, A.; Rayner, L.; Zhang, X.; Tuma, R.; Stockley, P.; Buck, M., Mutations in RNA Polymerase Bridge Helix and Switch Regions Affect Active-Site Networks and Transcript-Assisted Hydrolysis. *J Mol Biol* **2015**, *427* (22), 3516-26.
100. Weinzierl, R. O., The Bridge Helix of RNA polymerase acts as a central nanomechanical switchboard for coordinating catalysis and substrate movement. *Archaea* **2011**, *2011*, 608385.
101. Kireeva, M. L.; Domecq, C.; Coulombe, B.; Burton, Z. F.; Kashlev, M., Interaction of RNA polymerase II fork loop 2 with downstream non-template DNA regulates transcription elongation. *J Biol Chem* **2011**, *286* (35), 30898-910.
102. Naji, S.; Bertero, M. G.; Spitalny, P.; Cramer, P.; Thomm, M., Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res* **2008**, *36* (2), 676-87.
103. Mukhopadhyay, J.; Das, K.; Ismail, S.; Koppstein, D.; Jang, M.; Hudson, B.; Sarafianos, S.; Tuske, S.; Patel, J.; Jansen, R.; Irschik, H.; Arnold, E.; Ebright, R. H., The RNA polymerase "switch region" is a target for inhibitors. *Cell* **2008**, *135* (2), 295-307.
104. Boyaci, H.; Chen, J.; Lilic, M.; Palka, M.; Mooney, R. A.; Landick, R.; Darst, S. A.; Campbell, E. A., Fidaxomicin jams Mycobacterium tuberculosis RNA polymerase motions needed for initiation via RbpA contacts. *Elife* **2018**, *7*.
105. Buurman, E. T.; Foulk, M. A.; Gao, N.; Laganas, V. A.; McKinney, D. C.; Moustakas, D. T.; Rose, J. A.; Shapiro, A. B.; Fleming, P. R., Novel rapidly diversifiable antimicrobial RNA polymerase switch region inhibitors with confirmed mode of action in Haemophilus influenzae. *J Bacteriol* **2012**, *194* (20), 5504-12.
106. Molodtsov, V.; Fleming, P. R.; Eyermann, C. J.; Ferguson, A. D.; Foulk, M. A.; McKinney, D. C.; Masse, C. E.; Buurman, E. T.; Murakami, K. S., X-ray crystal structures of Escherichia coli RNA polymerase with switch region binding inhibitors enable rational design of squaramides with an improved fraction unbound to human plasma protein. *J Med Chem* **2015**, *58* (7), 3156-71.
107. Xu, M.; Zhou, Y. N.; Goldstein, B. P.; Jin, D. J., Cross-resistance of Escherichia coli RNA polymerases conferring rifampin resistance to different antibiotics. *J Bacteriol* **2005**, *187* (8), 2783-92.

108. Campbell, E. A.; Pavlova, O.; Zenkin, N.; Leon, F.; Irschik, H.; Jansen, R.; Severinov, K.; Darst, S. A., Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* **2005**, *24* (4), 674-82.
109. Zhang, Y.; Degen, D.; Ho, M. X.; Sineva, E.; Ebricht, K. Y.; Ebricht, Y. W.; Mekler, V.; Vahedian-Movahed, H.; Feng, Y.; Yin, R.; Tuske, S.; Irschik, H.; Jansen, R.; Maffioli, S.; Donadio, S.; Arnold, E.; Ebricht, R. H., GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *Elife* **2014**, *3*, e02450.
110. Maffioli, S. I.; Zhang, Y.; Degen, D.; Carzaniga, T.; Del Gatto, G.; Serina, S.; Monciardini, P.; Mazzetti, C.; Guglielame, P.; Candiani, G.; Chiriac, A. I.; Facchetti, G.; Kaltofen, P.; Sahl, H. G.; Deho, G.; Donadio, S.; Ebricht, R. H., Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **2017**, *169* (7), 1240-1248 e23.
111. Temiakov, D.; Zenkin, N.; Vassylyeva, M. N.; Perederina, A.; Tahirov, T. H.; Kashkina, E.; Savkina, M.; Zorov, S.; Nikiforov, V.; Igarashi, N.; Matsugaki, N.; Wakatsuki, S.; Severinov, K.; Vassylyev, D. G., Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol Cell* **2005**, *19* (5), 655-66.
112. Degen, D.; Feng, Y.; Zhang, Y.; Ebricht, K. Y.; Ebricht, Y. W.; Gigliotti, M.; Vahedian-Movahed, H.; Mandal, S.; Talaue, M.; Connell, N.; Arnold, E.; Fenical, W.; Ebricht, R. H., Transcription inhibition by the depsipeptide antibiotic salinamide A. *Elife* **2014**, *3*, e02451.
113. Bae, B.; Nayak, D.; Ray, A.; Mustaev, A.; Landick, R.; Darst, S. A., CBR antimicrobials inhibit RNA polymerase via at least two bridge-helix cap-mediated effects on nucleotide addition. *Proc Natl Acad Sci U S A* **2015**, *112* (31), E4178-87.
114. Lin, W.; Mandal, S.; Degen, D.; Liu, Y.; Ebricht, Y. W.; Li, S.; Feng, Y.; Zhang, Y.; Mandal, S.; Jiang, Y.; Liu, S.; Gigliotti, M.; Talaue, M.; Connell, N.; Das, K.; Arnold, E.; Ebricht, R. H., Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Mol Cell* **2017**, *66* (2), 169-179 e8.
115. Artsimovitch, I.; Svetlov, V.; Nemetski, S. M.; Epshtein, V.; Cardozo, T.; Nudler, E., Tagetitoxin inhibits RNA polymerase through trapping of the trigger loop. *J Biol Chem* **2011**, *286* (46), 40395-400.



116. Vassylyev, D. G.; Svetlov, V.; Vassylyeva, M. N.; Perederina, A.; Igarashi, N.; Matsugaki, N.; Wakatsuki, S.; Artsimovitch, I., Structural basis for transcription inhibition by tagetitoxin. *Nat Struct Mol Biol* **2005**, *12* (12), 1086-93.
117. Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A., Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *J Am Chem Soc* **2003**, *125* (41), 12475-83.
118. Walker, S. S.; Degen, D.; Nickbarg, E.; Carr, D.; Soriano, A.; Mandal, M.; Painter, R. E.; Sheth, P.; Xiao, L.; Sher, X.; Murgolo, N.; Su, J.; Olsen, D. B.; Ebright, R. H.; Young, K., Affinity Selection-Mass Spectrometry Identifies a Novel Antibacterial RNA Polymerase Inhibitor. *ACS Chem Biol* **2017**, *12* (5), 1346-1352.
119. Ma, C.; Yang, X.; Lewis, P. J., Bacterial Transcription Inhibitor of RNA Polymerase Holoenzyme Formation by Structure-Based Drug Design: From in Silico Screening to Validation. *ACS Infect Dis* **2016**, *2* (1), 39-46.
120. Andre, E.; Bastide, L.; Villain-Guillot, P.; Latouche, J.; Rouby, J.; Leonetti, J. P., A multiwell assay to isolate compounds inhibiting the assembly of the prokaryotic RNA polymerase. *Assay Drug Dev Technol* **2004**, *2* (6), 629-35.
121. Yang, X.; Ma, C.; Lewis, P. J., Identification of inhibitors of bacterial RNA polymerase. *Methods* **2015**, *86*, 45-50.
122. Skordalakes, E.; Brogan, A. P.; Park, B. S.; Kohn, H.; Berger, J. M., Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure* **2005**, *13* (1), 99-109.
123. Landick, R.; Krek, A.; Glickman, M. S.; Socci, N. D.; Stallings, C. L., Genome-Wide Mapping of the Distribution of CarD, RNAP sigma(A), and RNAP beta on the Mycobacterium smegmatis Chromosome using Chromatin Immunoprecipitation Sequencing. *Genom Data* **2014**, *2*, 110-113.
124. Srivastava, D. B.; Leon, K.; Osmundson, J.; Garner, A. L.; Weiss, L. A.; Westblade, L. F.; Glickman, M. S.; Landick, R.; Darst, S. A.; Stallings, C. L.; Campbell, E. A., Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* **2013**, *110* (31), 12619-24.

125. Stallings, C. L.; Stephanou, N. C.; Chu, L.; Hochschild, A.; Nickels, B. E.; Glickman, M. S., CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. *Cell* **2009**, *138* (1), 146-59.
126. Rammohan, J.; Ruiz Manzano, A.; Garner, A. L.; Stallings, C. L.; Galburt, E. A., CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Res* **2015**, *43* (6), 3272-85.
127. Davis, E.; Chen, J.; Leon, K.; Darst, S. A.; Campbell, E. A., Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* **2015**, *43* (1), 433-45.
128. Garner, A. L.; Weiss, L. A.; Manzano, A. R.; Galburt, E. A.; Stallings, C. L., CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. *Mol Microbiol* **2014**, *93* (4), 682-97.
129. Garner, A. L.; Rammohan, J.; Huynh, J. P.; Onder, L. M.; Chen, J.; Bae, B.; Jensen, D.; Weiss, L. A.; Manzano, A. R.; Darst, S. A.; Campbell, E. A.; Nickels, B. E.; Galburt, E. A.; Stallings, C. L., Effects of Increasing the Affinity of CarD for RNA Polymerase on Mycobacterium tuberculosis Growth, rRNA Transcription, and Virulence. *J Bacteriol* **2017**, *199* (4).
130. Weiss, L. A.; Harrison, P. G.; Nickels, B. E.; Glickman, M. S.; Campbell, E. A.; Darst, S. A.; Stallings, C. L., Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. *J Bacteriol* **2012**, *194* (20), 5621-31.
131. Gulten, G.; Sacchettini, J. C., Structure of the Mtb CarD/RNAP beta-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure* **2013**, *21* (10), 1859-69.

## Chapter II

### Construction, Expression, and Purification on *M. tuberculosis* RNA polymerase

MTB RNA polymerase (RNAP) is the target of rifampin (RMP), a first line treatment for TB. Rifampin binds to the  $\beta$ -subunit or rpoB of RNA polymerase (RNAP) preventing elongation of the nascent RNA transcript. The increased prevalence of rifampin resistance (RMP<sup>R</sup>) continues to make treatment options more challenging.<sup>1,2</sup> Three mutant RNAPs ( $\beta$ D435V,  $\beta$ H445Y, and  $\beta$ S450L) together represent roughly 80% of RMP<sup>R</sup> alleles found in clinical isolates, with the  $\beta$ S450L mutant representing the majority of mutants observed.<sup>3,4,5,6</sup> MTB RNAP is an ideal target for drug discovery because while sequence similarity among bacterial RNAPs is high, they structurally diverge from their eukaryotic counterparts. A single DNA-dependent RNAP is responsible for all RNA synthesis within bacteria. To study the therapeutic potential of RNAP for the treatment of TB, highly pure and active RNAP is required.

Several bacterial RNAPs have been successfully overexpressed and purified.<sup>7-12</sup> Most structural and biochemical studies use *Thermus aquaticus* (*Taq*) and *E. coli* RNAP, though other bacterial RNAPs have been successfully purified. *Taq* RNAP is useful for studies involving highly conserved motifs within bacterial RNAP such as catalytic center; however, there are significant sequence differences in regions of RNAP distal to the active center. *E. coli* RNAP has been purified and characterized for nearly 50 years and has become the gold standard for bacterial RNA polymerase studies involving drug discovery. Efficient expression and purification protocols are available for acquiring large

quantities of *E. coli* RNAP, the success of which is likely due to the homologous overexpression. Acquisition of other bacterial RNAPs via heterologous expression in *E. coli* has been more challenging.

For more thorough studies involving MTB RNAP function, the use of endogenous RNAP from MTB would be ideal. *E. coli* RNAP shares roughly ~40% sequence similarity with MTB RNAP (though the RMP binding cleft shares nearly 95% sequence similarity), so there are likely differences in both protein structure and function which may limit efforts to use *E. coli* in MTB drug discovery efforts. This has been shown to be especially true in the case of open-promoter complex studies.<sup>13</sup> Promoter complex studies show that MTB RNAP forms a much less stable open-promoter complex than the *E. coli* RNAP. Therefore, in order to improve the biochemical information about RNAP function, MTB RNAP should be used for mechanistic studies and drug discovery for TB therapeutics.

Several attempts to generate purified MTB RNAP have been attempted. These include production and isolation of MTB RNAP from MTB.<sup>14</sup> While purified enzyme was obtained using this method, it requires a laboratory meeting Biosafety Level 3 (BSL3) protocols. Additionally, MTB is a slow growing organism with a doubling time of nearly 24 hours so it takes several weeks to obtain cultures. Other methodologies employed to obtain MTB RNAP include overexpressing each of the subunits separately and reconstituting the RNAP *in vitro*.<sup>15</sup> The issue with this method is a significant amount of the reconstituted protein was inactive, likely due to aggregation, misfolding of purified subunits, and misassembly of RNAP complex.

Herein, efforts to obtain pure and active MTB RNAP are discussed. A method previously developed to overexpress MTB from a polycistronic vector encoding the core

subunits of MTB RNAP are discussed.<sup>16</sup> A newly developed method, where multiple plasmids encoding multiple MTB core subunits as well as the primary sigma factor, SigA, are also described.<sup>17</sup> The effect of the RMP<sup>R</sup> mutations in the RRDR on RMP potency are also explored.

## Materials and Methods

### *Plasmid and DNA Manipulation*

pMTBRP-5,6,7,8 were all constructed as previously described.<sup>16</sup> MTB RNAP Duet-expression vectors were constructed as follows. Genes encoding RpoB ( $\beta$ -subunit) and RpoC ( $\beta'$ -subunit) were amplified for construction of pET-rpoB(C-terminal His<sub>10X</sub>)-rpoC, containing a C-terminal decahistadine-tag on the RpoB gene, from pMTBRP-5 using primers rpoB AMP FOR and rpoB AMP REV and rpoC AMP FOR and rpoC AMP REV respectively ( all construction primer are available in **Appendix Table II-A1**). For RMP<sup>R</sup> mutants  $\beta$ D435V,  $\beta$ H445Y, and  $\beta$ S450L the *rpoB* genes were amplified from pMTBRP-6, pMTBRP-7, and pMTBRP-8 respectively using the primers mentioned above for the amplification of the WT *rpoB* gene. Amplified *rpoB* was digested with NcoI and BamHI and *rpoC* was digested with NdeI and EcoRV and subcloned into alkaline phosphatase (CIP) treated pET-Duet which was digested with the corresponding digestive enzymes. The *rpoB* gene was subcloned first into pET Duet followed by *rpoC*.

pAcYc-rpoA-rpoZ, encoding RpoA and RpoZ, was constructed by amplifying *rpoA* and *rpoZ* from pMTBRP-5 using rpoA AMP FOR and rpoA AMP REV and rpoZ AMP FOR and rpoZ AMP REV. The same methodology which was used for the construction of pET-rpoB(C-terminal His<sub>10X</sub>)-rpoC was used for pAcYc-rpoA-rpoZ. pRSF-SigA, encoding MTB SigA, was produced by amplifying *sigA* from H37Rv gDNA using SigA AMP FOR and

SigA AMP REV, restriction digest of the amplified DNA, and ligation into CIP treated pRSF using the NcoI and MfeI restriction sites. All DNA fragments and digested vectors were gel purified on agarose gels and extracted with Qiagen Gel Extraction kit as directed by the manufacturer prior to each ligation step.

To produce the pET-rpoB-rpoC, which contains no decahistidine-tag the *rpoB* gene was removed by restriction digest and replaced by the *rpoB* gene which does not contain the histidine affinity tag. The *rpoB* gene was amplified using rpoB AMP 2 FOR and rpoB AMP 2 REV. A hexahistidine-tag was inserted at the N-terminus of the *rpoA* gene by amplifying the *rpoA* gene with rpoA FOR NThis and rpoA AMP REV from pAcYc-rpoA-rpoZ, digesting the DNA fragment with BamHI and HindIII, and ligating into CIP, BamHI, and HindIII treated pAcYc-rpoZ.

To produce pAcYc-rpoA(N-terminal His<sub>10X</sub>)-SigA, *rpoA* was amplified with rpoA N10his FOR and rpoA AMP REV, digested with NcoI and HindIII, and ligated into pAcYc which was treated with NcoI, HindIII, and CIP. The *sigA* gene was amplified from pRSF-SigA with SigA NdeI AMP FOR and SigA KpnI AMP REV and treated with NdeI and KpnI. Digested DNA fragment was ligated in CIP and restricted pAcYc-rpoA(N-terminal His<sub>10X</sub>). pRSF-rpoZ was produced by amplifying *rpoZ* from pAcYc-rpoA-rpoZ with primers rpoZ NcoI AMP FOR and rpoZ MfeI AMP REV and processed in the same fashion as pRSF-SigA mentioned above. All plasmids mentioned above were sequenced at the University of Michigan Sequencing core.

#### *Protein expression and purification*

Initial attempts of expression and purification of MTB RNAP core from pMTBRP-5,6,7,8 were performed as previously described.<sup>16</sup> Expression and purification of MTB

RNAP core and holoenzyme using the Duet expression system were performed using the same methodology used for the pMTBRP expression system with the following difference. For acquiring core MTB RNAP pET-rpoB(His<sub>10X</sub>)-rpoC and pAcYc-rpoA-rpoZ, plasmids were transformed into BL21(DE3) chemically competent *E. coli* in the presence of carbenicillin and chloramphenicol. When holoenzyme was produced both previously mentioned plasmids, in addition to pRSF-SigA, were transformed into BL21(DE3) chemically competent *E. coli* and grown to an OD<sub>600</sub> of 0.4 before cooling to room temperature and inducing with 0.5mM IPTG for 16 hours and 16°C. Separation techniques for early attempts to acquire purified protein utilized immobilized metal affinity chromatography (IMAC) and gel filtration (see **Appendix Figure II-A1**).

Final protocols for the purification *E. coli* and MTB RNAP holoenzymes used for all *in vitro* biochemical assays were prepared as follows. WT and RIF<sup>R</sup> MTB RNAP holoenzymes were overexpressed in BL21(DE3) cells transformed with the three plasmids: pET-rpoB-rpoC, pAcYc-rpoA(N-terminal His<sub>10X</sub>)-SigA, and pRSF-rpoZ. *E. coli* RNAP holoenzymes were overexpressed from the plasmid pVS10 and pRSF-Sig70 in terrific broth supplemented with 1 mM ZnSO<sub>4</sub> and induced as described in Banerjee et al,<sup>17</sup> with the addition of a second plasmid, pRFS-Sig70 encoding rpoD. BL21(DE3) cells transformed with the WT or RIF<sup>R</sup> MTB RNAP holoenzyme expression vectors were grown to an OD<sub>600</sub> of 0.4 in terrific broth supplemented with 1 mM ZnSO<sub>4</sub> and induced with 0.5 mM IPTG for 16 hours at 16 °C. Cell pellets from 2L of culture were resuspended in 20 mL lysis buffer (20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 μM ZnCl<sub>2</sub>, 5% glycerol, 2 mM β-mercaptoethanol, 1 mM PMSF and 1X Roche cOmplete ULTRA protease cocktail). Cells were lysed by sonication and clarified by centrifugation at 25,000xg for 45 min. The

RNAP•DNA complexes were precipitated by addition of polyethyleneimine (pH 7.9) to 0.6% followed by centrifugation at 6,000xg for 10 min. The pellet was resuspended in 12 mL lysis buffer containing 500 mM NaCl, followed by centrifugation at 6,000xg for 10 min. After centrifugation again and discard of the supernatant, the RNAP was solubilized in 20 mL lysis buffer containing 1 M NaCl. The RNAP was precipitated by adding ammonium sulfate (0.3 g/mL of solution) followed by a 60 min incubation at 4°C. The RNAP precipitate was pelleted by centrifugation (6,000xg for 10 min) and the ammonium sulfate pellet was resuspended, sterile filtered, and dialyzed overnight (20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 20  $\mu$ M ZnCl<sub>2</sub>, 5% glycerol, 2 mM  $\beta$ -mercaptoethanol).

The dialyzed RNAP solution was applied to a 5 mL Source 15S cation exchange column (GE Healthcare) in TGEB (10 mM Tris-HCl (pH 8.0), 5% glycerol, 0.1 mM EDTA, 5 mM  $\beta$ -mercaptoethanol) with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl. Isolated holoenzyme was then further purified on a 5 mL HisTrap HP column. RNAP was applied to the column in 10 mM Tris-HCl (pH8.0), 20 mM NaCl, 5% glycerol, 5 mM  $\beta$ -mercaptoethanol and eluted over a linear gradient to 500 mM imidazole. The peak containing holoenzyme was then applied to a 5 mL Source 15Q anion exchange column (GE Healthcare) in TGEB with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl. Purified RNAP holoenzyme was dialyzed into storage buffer (40 mM Tris-HCl (pH 7.9), 200 mM NaCl, 0.1 mM EDTA, 1 mM MgCl<sub>2</sub>, 20  $\mu$ M ZnCl<sub>2</sub>, 50% glycerol, and 5 mM DTT) and stored at -20°C.



### *Primer extension Assays*

Determination of the concentration of active RNAP was conducted using a DNA•RNA scaffold that mimics the transcription elongation complex.<sup>18</sup> First, an 8 nucleotide RNA (IDT technologies) was labeled at the 5' end with [ $\gamma$ -P<sup>32</sup> GTP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs). Next, the T4 polynucleotide kinase was heat inactivated. The DNA•RNA duplex scaffolds were prepared by mixing equimolar amounts of template DNA (5'-CTGAAATCGACATCGCCGCTCAACAT-3') and 5'-labeled RNA (5'-GCGGCGAU-3'), heating to 85 °C for 10 min, followed by slow cooling to room temperature. The elongation complex was formed by mixing equimolar RNAP with the DNA•RNA duplex scaffold in reaction buffer at 28 °C for 20 min. The reaction buffer for the *E. coli* RNAP consisted of 40 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM MgCl<sub>2</sub>, 8 mM DTT, 0.01% Triton X-100 and for the MTB RNAP was 40 mM Tris-HCl (pH 7.9), 40 mM potassium glutamate, 0.1 mM EDTA, 10 mM MgCl<sub>2</sub>, 25 µg/mL BSA, 1 mM DTT; both reactions contained 15 mU of inorganic pyrophosphatase (New England Biolabs). The primer extension reaction was conducted for 10 min after adding the next incoming GTP (400 µM final concentration) to the elongation complex. RNA extension was stopped by adding an equal volume of 2x RNA loading buffer (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.025% ethidium bromide, 0.5 mM EDTA – Thermo Scientific). RNAs were resolved by gel-electrophoresis in 20% acrylamide (19:1), 7 M urea gel with 1x TBE. The gels were processed using a PhosphorImaging cassette (Typhoon 8600 scanner) and the densitometry of the RNA bands were quantified using ImageQuant software (GE Healthcare).

## *In Vitro* Transcription Assays

An *in vitro* plasmid based transcription assay using a Malachite Green Aptamer (MGA) gene that was previously developed for high-throughput screening was modified for these studies.<sup>19</sup> Individual modified pTZ18U vectors containing the *E. coli* *rrnB* P1 (-66 to +5) and MTB *rrnA* P3 (-55 to +15) ribosomal RNA promoters followed by 4 consecutive repeats of DNA encoding the MGA and three consecutive repeats of the *synB* artificial terminator sequence were prepared, pMGA4-Ec-*rrnB*1-SynBx3 and pMGA4-Mt-*rrnA*3-SynBx3, respectively. The *synB* terminator was shown to terminate transcription with a 88% and 98% termination efficiency for *E. coli* and *M. bovis* RNAPs, respectively<sup>20</sup>.

The buffer for determining the RMP IC<sub>50</sub>s for the *E. coli* RNAP holoenzymes was: 20 nM pMGA4-Ec-*rrnB*1-SynBx3, 40 mM Tris-HCl (7.5 at 22 °C), 50 mM KCl, 10 mM MgCl<sub>2</sub>, 0.01% Triton X-100, 1 mM DTT, 500 μM each NTP. For MTB RNAP holoenzymes, 50 nM pMGA4-Mt-*rrnA*3-SynBx3, 40 mM Tris-HCl (pH 8.0 at 37°C), 150 mM potassium glutamate, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 25 μg/mL BSA, 1 mM DTT, 500 μM each NTP was used. For WT *E. coli* and MTB RNAPs, 10 nM of holoenzyme was used, and for RMP<sup>R</sup> *E. coli* and MTB RNAPs, 100 nM enzyme was used. In all cases a 3-fold excess of the corresponding purified σ factor was also added. RMP added to the assays was dissolved in DMSO; therefore, DMSO was added to a final concentration of 4% (which had a negligible effect on the assay) in all assays.<sup>19</sup> RMP was allowed to incubate with holoenzyme for 30 min prior to the addition of template DNA and NTPs. Reactions were incubated at 37°C for 90 min, followed by quenching on ice with addition of ice cold malachite green (MG) in water to a final concentration of 75 μM MG. Fluorescence was

detected at excitation and emission wavelengths of 628 nm and 660 nm, respectively, using a BioTek Synergy H1 plate reader. The fluorescence readings were normalized to % activity, plotted against the log [RMP] and fit by nonlinear regression to Eq. (1) where  $M_0$  = log of compound concentration,  $M_1$  = log of  $IC_{50}$ ,  $M_2$  = hill slope,  $M_3$  = lower limit of the curve, and  $M_4$  = upper limit of the curve. Three replicate sets of data for each RNAP were individually fit and the averages and standard deviations of the three individual  $IC_{50}$ s are reported in Table 1. For clarity of visualization, the average % activity for each [RMP] were plotted, with their standard deviations as error bars. The final concentrations of the WT RNAPs were 10 nM, therefore the minimum  $IC_{50}$  which can be determined for both *E. coli* and MTB WT RNAPs was 5 nM.

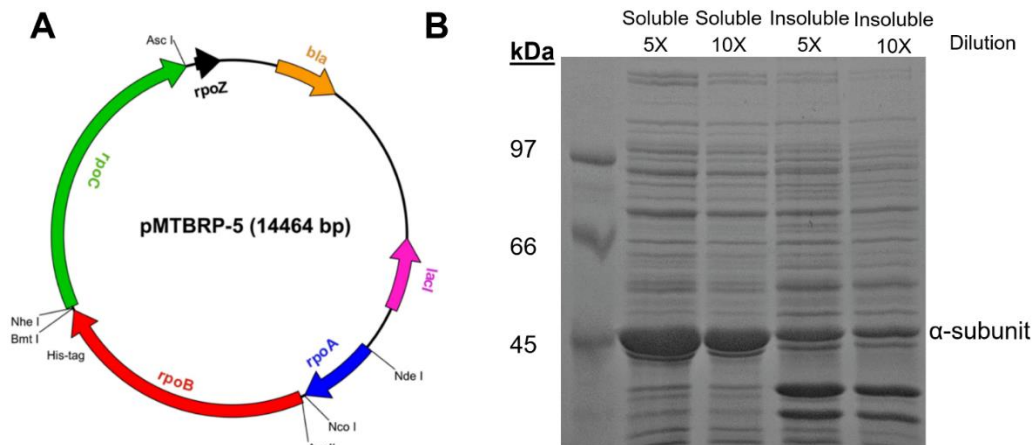
$$(1) y = M_3 + \frac{(M_4 - M_3)}{(1 + 10^{(M_0 - M_1) \times M_2})}$$

To determine contamination by chromosomally-encoded WT *E. coli* RNAP, transcription assays for the MTB and *E. coli* RMP<sup>R</sup> RNAP mutants were conducted in the presence and the absence of 200 nM RMP. The results indicate that such contamination in MTB RNAP preparations is negligible, while ~25% of *E. coli* RIF<sup>R</sup> RNAP preparations were contaminated with endogenous WT RpoB (see **Fig. 3B**).

## Results

### *Preparation and characterization of MTB RNAPs from pMTBRP-5*

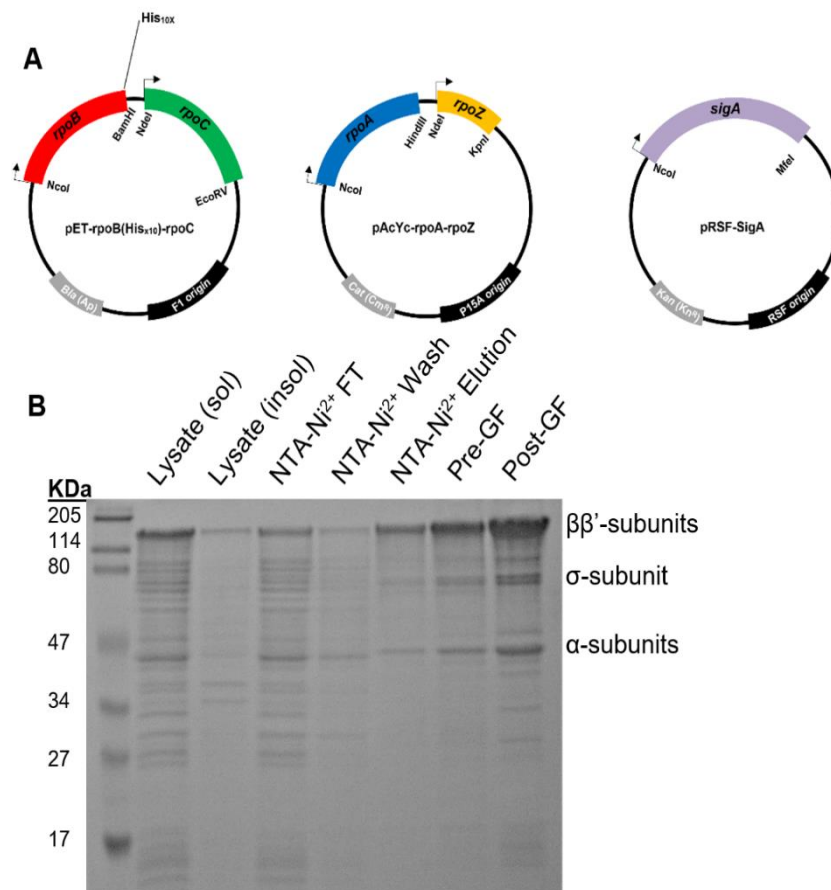
Initial attempts at acquiring purified MTB RNAP used a single expression vector pMTBRP-5, under control of a single T7 promoter upstream of the  $\alpha$ -subunit or RpoA (see **Figure II-1A**). The system produces a >15 kilobase polycistronic mRNA which encodes all subunits of core RNAP.<sup>21</sup> The only overexpressed MTB RNAP subunit which could visually identified was the  $\alpha$ -subunit (see **Figure II-1B**). Enrichment of target protein using IMAC chromatography resulted in low yield and samples with poor purity. Several methods were explored to improve overexpression of genes downstream of *rpoA*. These include optimization of the ribosome binding sites (RBS), removal of rare codons, and insertion of additional T7 promoter/lac operon sites upstream of each subunit.



**Figure II-1: Overexpression of MTB RNAP with pMTBRP. A)** pMTBRP plasmid map. **B)** SDS-PAGE of overexpression of MTB RNAP from the pMTBRP system.

## Preparation of MTB RNAP from the Duet Expression System

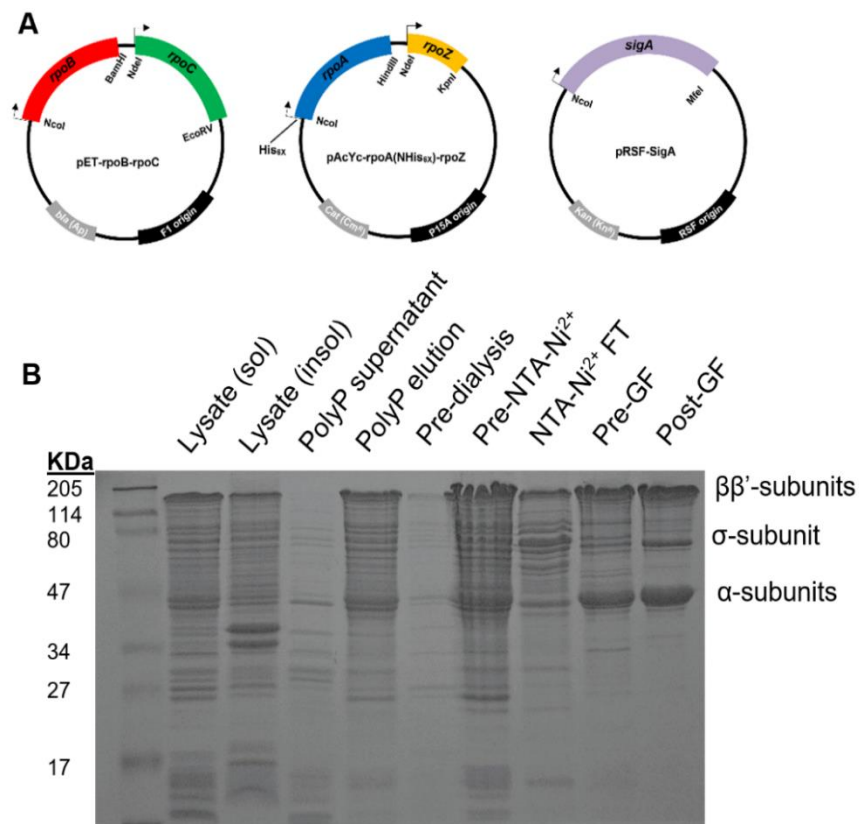
For preparing WT and RMP<sup>R</sup> mutant MTB RNAPs, we used an overexpression system in a *E. coli* host strain<sup>16-17</sup>, which utilize vectors carrying the genes of the RNAP subunits, under control of a strong T7 promoter. Our first generation of MTB expression vectors expressed the RpoB and RpoC on a single plasmid (pET) and RpoA and RpoZ on a second plasmid (pAcYc). When holoenzyme was desired a third plasmid (pRSF) containing MTB SigA was added in addition to the pET and pAcYc vectors. A decahistidine affinity tag was present on the C-terminus of RpoB (see **Figure II-2A**).



**Figure II-2: Purification of MTB RNAP with 1<sup>st</sup>-generation Duet Expression System.** **A)** The three co-expression vectors used to produce MTB holoenzyme with a C-terminal decahistidine tag on the β-subunit. **B)** SDS-PAGE of purification methodology used to acquire MTB RNAP holoenzyme.

Overexpression of all MTB RNAP subunits can be observed in induced cell lysates with the exception of the  $\omega$ -subunit. MTB RNAP could also be immobilized, washed, and eluted using NTA-Ni<sup>2+</sup> agarose resin. As a final polishing step MTB RNAP is further purified using gel filtration (GF) on a HiPrep Sephacryl S-200 HR column. Though a significant amount of MTB RNAP was obtained using this methodology (2-5 mg of protein per L culture) a significant amount of contaminating proteins remained present after the final purification step (see **Figure II-2B**). Additionally, the  $\omega$ -subunit was not observed in the final preparation. In *E. coli* RNAP crystal structures the N- and C-terminus of the  $\beta$ - and  $\beta'$ -subunit coalesce in proximity to the site where the  $\omega$ -subunit associates with RNAP. Concerns that the His-tag on the C-terminus of the  $\beta$ -subunit could be interfering prompted switching the location of the His-tag to the N-terminus of the  $\alpha$ -subunit (see **Figure II-3A**).

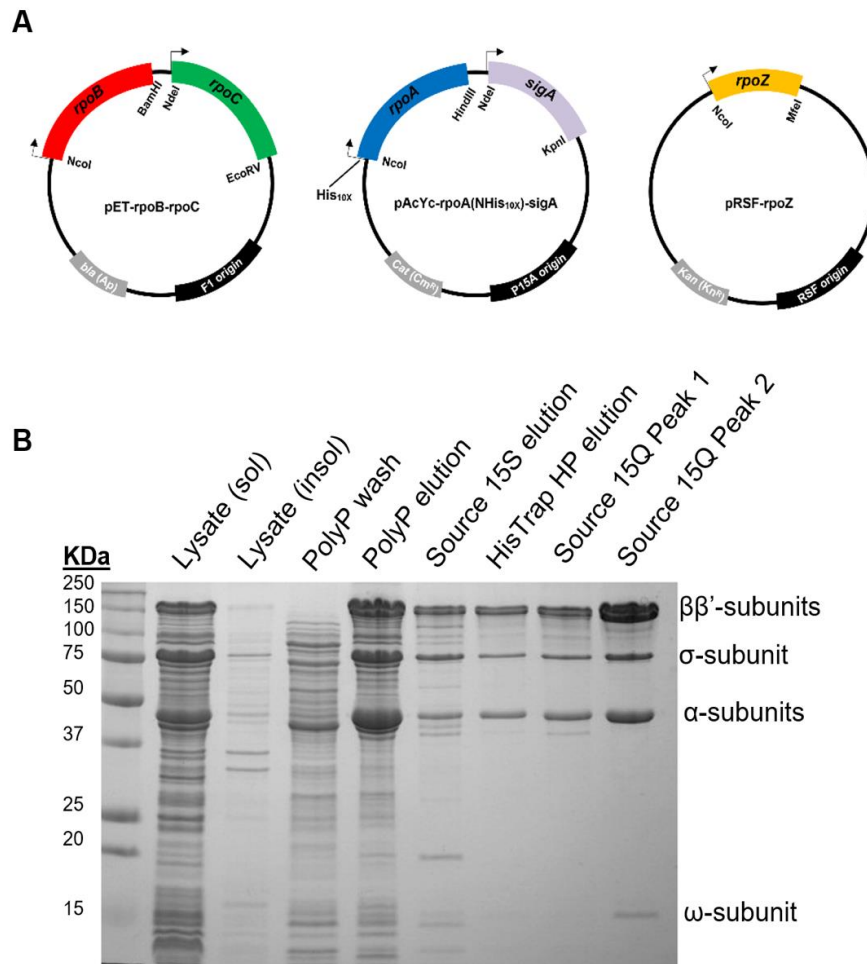
Our co-expression system was modified by removal of the histidine tag from the C-terminus of *rpoB* on pET-*rpoB*(His<sub>x10</sub>)-*rpoC* to make pET-*rpoB*-*rpoC* and addition of a N-terminal hexahistidine tag on the *rpoA* to make pAcYc-*rpoA*(His<sub>6x</sub>)-*rpoZ*. Switching the placement of the histidine tag resulted a large stoichiometric excess of  $\alpha$ -subunit in final preparation. Again, the  $\omega$ -subunit was not observed in final preparations. Contaminating proteins were also present in the final purified protein (see **Figure II-3B**).



**Figure II-3: Purification of MTB RNAP with 2<sup>nd</sup>-generation Duet Expression System.**  
**A)** The three co-expression vectors used to produce MTB holoenzyme with a N-terminal hexahistidine tag on the  $\alpha$ -subunit. **B)** SDS-PAGE of purification methodology used to acquire MTB RNAP holoenzyme.

The 3<sup>rd</sup> generation of co-expression vectors used pETrpoB-rpoC; however, the *sigA* gene was exchanged with *rpoZ* and the hexahistidine tag on *rpoA* was extended to a decahistidine tag (see **Figure II-4A**). Additionally, our purification protocol was significantly altered to improve the quality and stoichiometry of the final purified product (see **Figure II-4B**, **Appendix Figure II-A1**). The polyamine P precipitation and ammonium sulfate steps were retained in the final purification method. Since the  $\alpha$ -subunit was overexpressed in such high quantities, the first chromatography step was cation exchange chromatography (see **Appendix Figure II-A2**). RNAP holoenzyme is retained on the Source 15S column, allowing core and excess  $\alpha$ -subunit to be removed

from the sample. The second chromatography step was IMAC using a HisTrap HP column (see **Appendix Figure II-A3**). This led to further enrichment of MTB RNAP and removed endogenous *E. coli* RNAP. The last polishing step used anion exchange chromatography. MTB RNAP eluted as two peaks from the Source 15Q column, the second peak had a higher activity and the  $\omega$ -subunit was also present in the sample (see **Appendix Figure II-A4**). The stoichiometry of the subunits was also similar to the theoretical values. The resulting protocol produced highly pure RNAP with a high level of catalytic activity.

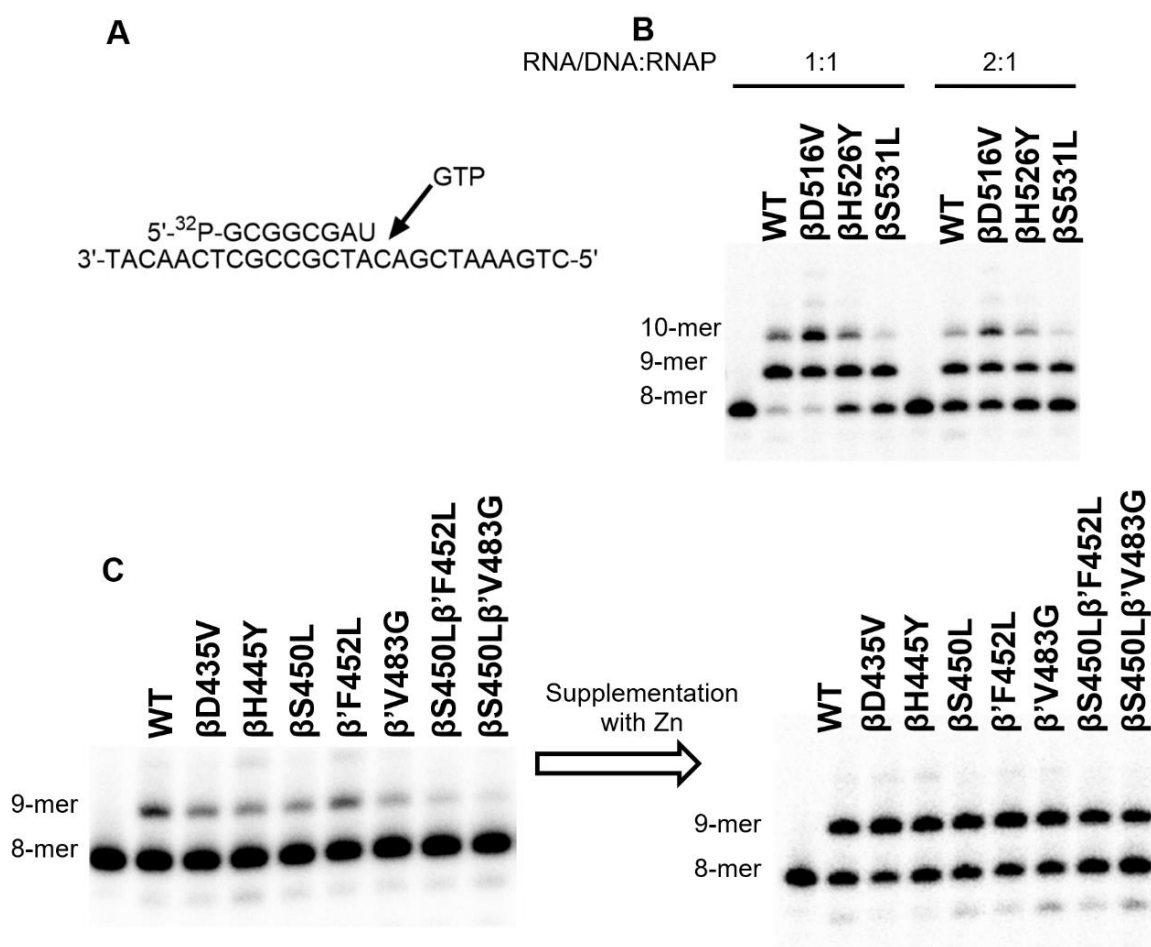


**Figure II-4: Purification of MTB RNAP with 3<sup>rd</sup>-generation Duet Expression System.** **A)** The three co-expression vectors used to produce MTB holoenzyme with a C-terminal decahistadine tag on the  $\beta$ -subunit. **B)** SDS-PAGE of purification methodology used to acquire MTB RNAP holoenzyme.



### Determination of percent of catalytically active RNAP from purifications

To determine the percent of active RNAP complex in a given preparation, a single nucleotide turnover assay using an elongation complex mimetic was used (see **Figure II-5**).<sup>18</sup> Initial purification methods produced RNAP with <10% catalytically competent complex. Literature which purported that a zinc-binding domain in the  $\beta'$ -subunit played a significant role in protein folding and assembly led to supplementation of growth media with ZnSO<sub>4</sub>.<sup>22</sup> MTB RNAP holoenzyme produced from these cultures ranged from 40-70% active complex. *E. coli* RNAPs were 80-100% active.



**Figure II-5: Determination of percent active complex via primer extension assay. A)** 5'-<sup>32</sup>P-RNA/DNA scaffold used for primer extension experiments. **B)** Activity of WT and RMP<sup>R</sup> *E. coli* RNAPs. **C)** Activity of WT and RMP<sup>R</sup> MTB RNAPs before (left) and after (right) supplementation of growth media with ZnSO<sub>4</sub>.

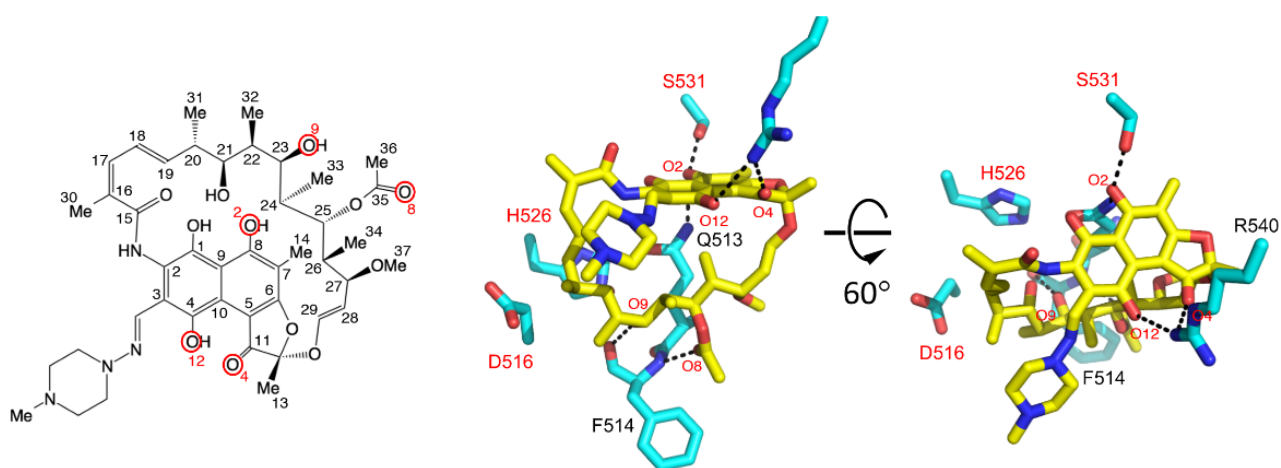
### *Determination of IC<sub>50</sub> of Rifampin with MTB and E. coli, WT and RMP<sup>R</sup> RNAPs*

*In vitro* characterization of both MTB and *E. coli* RNAPs were performed in a fluorescence-based plasmid assay. Two plasmids were used to characterize each organism's respective RNAP, one with the MTB *rrnAP3* promoter and the other with the *E. coli* *rrnBP1* promoter. Downstream of each promoter is four repeats of a malachite aptamer sequence which, when transcribed, adopts a secondary structure which has high binding affinity for malachite green. Binding of malachite green to aptamer prevents vibrational deexcitation of malachite green resulting in a >2000-fold increase in fluorescence. Downstream of the malachite green aptamer coding region, there are three consecutive synthetic terminators (*synB*).<sup>20</sup> These terminators prevent RNAP from circling the entire plasmid, which could potentially circumvent identification of inhibitor effects which occur during initiation.

The transcriptional activities of the WT RNAPs from *E. coli* and MTB are completely inhibited with 200 nM RMP, whereas the activities of RMP<sup>R</sup> RNAPs are not influenced by RMP at this concentration. Transcription assays of the RMP<sup>R</sup> RNAP mutants in the presence and the absence of 200 nM RMP indicated that contamination by chromosomally-encoded *E. coli* RNAP in the purified MTB RMP<sup>R</sup> RNAPs is negligible, while the *E. coli* RMP<sup>R</sup> RNAPs had ~25% contamination (see **Appendix Figure II-A4**).

The effect of RNAP mutation on the RMP sensitivity was evaluated by determining the IC<sub>50</sub> values for RMP using a promoter-dependent transcription assay with RNAP holoenzymes (*E. coli*  $\sigma^{70}$  holoenzyme and MTB  $\sigma^A$  holoenzyme). Three RMP<sup>R</sup> mutants tested in this study are classified into two groups according to their RMP sensitivities; the D516V and S531L mutants of *E. coli* RNAP are less sensitive to RMP (IC<sub>50</sub> 398 and 263

$\mu\text{M}$ , respectively), whereas the H526Y mutation makes this enzyme essentially insensitive to RMP ( $\text{IC}_{50} \geq 2 \text{ mM}$ ). Corresponding RMP<sup>R</sup> mutations in MTB RNAP were also classified into the same groups; the D435V and S450L mutants are less sensitive to RMP ( $\text{IC}_{50}$  880 and 789  $\mu\text{M}$ , respectively), whereas the H445Y mutant is essentially insensitive to RMP ( $\text{IC}_{50} > 2 \text{ mM}$ , or the solubility limit)<sup>23</sup> (see **Figure II-6, Appendix Figure II-A5**). These results support the use of the *E. coli* RNAP as a model of the MTB RNAP in probing molecular mechanisms of RIF resistance.



<i>E. coli</i> RNAPs	WT	D516V	H526Y	S531L
RMP $\text{IC}_{50}$ ( $\mu\text{M}$ )	< 0.005	398 ( $\pm$ 118)	> 2000	263 ( $\pm$ 26)
MTB RNAPs	WT	D435V	H445Y	S450L
RMP $\text{IC}_{50}$ ( $\mu\text{M}$ )	< 0.005	880 ( $\pm$ 176)	> 2000	789 ( $\pm$ 249)

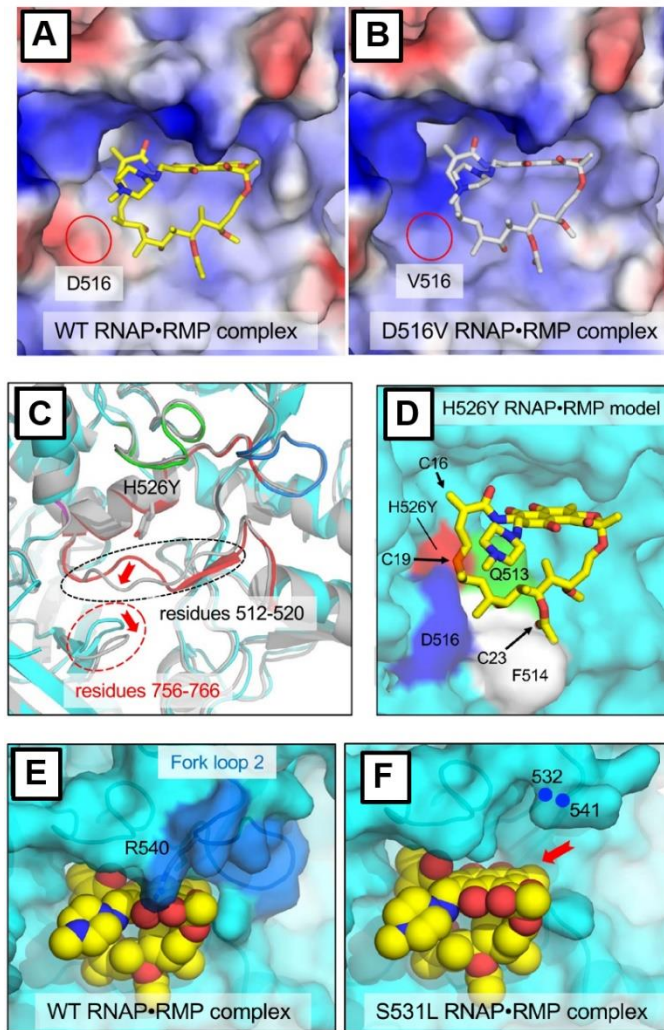
**Figure II-6: Inhibition of WT and RMP<sup>R</sup> MTB RNAPs by rifampin. A)** Structure of rifampin with key sites of interaction with RMP<sup>R</sup> mutations highlighted. **B)** RNAP  $\text{IC}_{50}$  values for rifampin.

### *Structural basis for resistance conferred by the Rif<sup>R</sup> Mutations in E. coli*

In WT RNAP and the RNAP•RMP complex, the D516 (D435 in MTB) residue forms part of the sidewall of the RMP binding pocket and assists in positioning RMP to form two hydrogen bonds between the F514 main chain (both amino and carboxyl groups) and the RMP ansa bridge (OH at C23, O9 atom and keto oxygen at C35, O8 atom) (see **Figure II-6A**). The D516 side chain also contributes to RMP binding by van der Waals interactions with the RMP ansa bridge around position C20 and/or by charge neutralization of basic residues (R529 and R687) in the RIF binding site to facilitate binding the relatively apolar RMP (see **Figure II-6A**). There is no major structural change in RRDR cluster I in the D516V mutant. The structure of the D516V RNAP•RMP complex is also similar to the WT RNAP•RMP complex. A significant difference between the WT and D516V RNAPs is the electrostatic distribution of the RMP binding pocket around C17 to C20 of the RMP ansa bridge, which becomes more basic due to losing the Asp side chain (see **Figure 7-IIA, B**). The altered electrostatic distribution of the RIF binding pocket may make it less favorable for binding the relatively apolar RMP.

In the WT RNAP•RMP complex, the H526 (H445 in MTB) residue is involved in forming the back wall of the RIF binding cavity and may form a hydrogen bond with oxygen atoms of the ansa bridge and/or the Q513 side chain, the latter of which defines the RMP binding wall within RRDR cluster I. We determined the crystal structure of the H526Y RNAP mutant at 3.6 Å resolution. The structure showed large rearrangements of main chain around the RRDR (residues 512-520 and 756-766) due to the H526Y substitution, which alters the shape of RMP binding interface (see **Figure II-7C, D**). Due to its very poor affinity for RMP, we could not obtain a crystal structure of the H526Y

RNAP•RMP complex. Modeling RMP from the WT RNAP•RMP complex into the H526Y mutant RNAP RMP-binding pocket shows that the rearranged RMP binding pocket would sterically clash with the rigid plane of the hydrophobic ansa-bridge (from C16 to C23 positions), particularly at C19 position (see **Figure II-7B**), therefore effectively preventing RMP binding and making the mutant virtually insensitive to RMP (**Table 1**).



**Figure II-7: Structural basis for inhibition of rifampin by RMP<sup>R</sup> mutant *E. coli* RNAPs.** **A)** Electrostatic surfaces of the RMP binding pocket of the WT and **B)** D516V mutant RNAPs complexed with RMP (stick models). RNAP surfaces are colored with positive (blue), negative (red) and neutral (white) electrostatic potentials. Positions of the D516 residue in WT RNAP (**A**) and V516 residue in the mutant (**B**) are indicated by red circles. **C)** Comparison of the WT and H526Y mutant RNAPs. RNAP structures are depicted as ribbon models and were superposed at their RRDRs. **D)** Steric hindrance of RMP binding to the H526Y mutant (left: the WT RNAP•RMP complex; right: the H526Y RNAP•RMP complex model). The locations of the clash between the ansa bridge of RMP with H526Y are indicated by arrows. **E)** The RMP RIF binding sites of the WT RNAP•RMP has fork loop 2 ordered upon RMP binding. **F)** Fork loop 2 is disordered in the S531L RNAP•RMP complex. Area of the RMP naphthalene ring exposure to solvent due to disordering of fork loop 2 is indicated by a red arrow.

In the WT RNAP•RMP complex, the S531 (S450 in MTB) residue is located in the deep cleft of the RMP binding pocket and forms a hydrogen bond with the phenol on C8 (O2 atom) of the RMP naphthalene ring (see **Figure II-6A**) and the top of the naphthalene ring is covered by  $\beta$  subunit loops, including fork loop 2 (residues 534-541) <sup>24-26</sup>.

In the crystal structure of S531L RNAP•RMP complex, however, a major difference is observed around the  $\beta$  subunit fork loop 2 of which the electron density is weak and scattered compared to its counterpart in the WT RNAP•RMP complex. Consequently, about half of the RMP naphthalene ring is exposed to solvent (see **Figure II-7E, F**), reducing ~40% of the contact area between RNAP and RMP (386 Å<sup>2</sup> and 244 Å<sup>2</sup> in the WT and S531L mutant, respectively). This alteration dramatically reduces the van der Waals interactions resulting in a large loss of binding free energy and is consistent with the ~10<sup>5</sup>-fold increase in IC<sub>50</sub> of RMP for the S531L mutant compared to WT RNAP. Interestingly, fork loop 2 is ordered in the S531L mutant without RMP bound (**Figure II-7E**), suggesting that the collision between RMP and the Leu side chain of the S531L mutant pushes fork loop 2 away, causing the observed disorder.

## Discussion

In order to characterize MTB RNAP, highly pure and active RNAP is required. Purification methods of bacterial RNAPs have been widely reported for *E. coli*, but heterologous overexpression of MTB RNAP remained a challenge.<sup>14-16</sup> Methodologies which used live MTB to produce protein were not a viable option due to safety concerns as well as the time needed to grow cultures and isolate protein.<sup>14</sup> The efficacy of reconstitution methods is low due to improper folding of protein.<sup>15</sup> *In vivo* reconstitution of RNAP in the standard *E. coli* host system represents the most viable option for acquiring MTB RNAP.

Initial attempts of MTB purification using the polycistronic vector pMTBRP series, were inefficient due to the disproportionate representation of MTB RNAP subunits which were overexpressed. Though the exact cause of this defect was not elucidated several attempts to identify the cause were explored. pMTBRP has a single T7 promoter, T7 RNAP is a highly processive polymerase capable of producing transcripts >15 kilobases, more than enough to cover the coding region on pMTBRP.<sup>27</sup> Several theories were produced to explain why the  $\alpha$ -subunit was the only observable subunit overexpressed in cell lysates (see **Figure II-1**). Identification of several rare codons, including the AGG and AGA codon (coding for Arg) each of which have an abundance of 3% and 4% can lead to pausing events during translation due to a low abundance of the corresponding amino acid charged tRNAs.<sup>28</sup> Additionally, these sequences, particularly if clustered, are similar to the shine-delgarno sequence which can lead to frameshifting along the template and eventual premature termination of translation.<sup>29</sup> Efforts to remove several of these rare codon sites from *rpoC* were ineffective at recovering overexpression of RpoC protein. As



attempts to introduce additional T7/lacO cis elements upstream of each of the subunits was underway a new methodology of overexpression of MTB RNAP was published.<sup>17</sup> This new method uses the Duet expression system. Since the method demonstrated production of all subunits efforts to identify the defect with the pMTBRP system were ended.

Banerjee published an optimized method for overexpression of MTB RNAP using the Duet expression system.<sup>17</sup> In this method two MTB RNAP subunits are introduced into each vector, each of which has two independent multiple cloning sites each under control of a T7/lac operator. Using this method, the yield of MTB RNAP was reported to be ~60 mgs of protein per liter of culture.<sup>17</sup> In his system, the decahistidine tag was placed on the N-terminus of the  $\alpha$ -subunit. However, since we were going to be studying RMP<sup>R</sup> mutations in the  $\beta$ -subunit, we decided to place the decahistidine on the C-terminus of the  $\beta$ -subunit (see **Figure II-2A**). MTB subunits from the pMTBRP expression system were subcloned into pET Duet, pAcYc Duet, and pRSF all of which can be co-expressed in *E. coli* because each has a unique origin of replication. Initial trials demonstrated that all MTB RNAP subunits could be overexpressed (though the  $\omega$ -subunit could not be distinguished in crude lysates, see **Figure II-2B**). Purification of MTB RNAP using this new system yielded roughly 2-5 mg of protein per liter of culture, significantly less than the 60 mg of protein reported by Banerjee. Additionally, purified MTB RNAP appeared to lack omega subunit in final preparations. Our observation that the N- and C-termini of the  $\beta$ - and  $\beta'$ -subunit coalesce in proximity to where the  $\omega$ -subunit associates with RNAP prompted moving the histidine-tag from the C-terminus of the  $\beta$ -subunit to the N-terminus of the  $\alpha$ -subunit.

In our 2<sup>nd</sup> generation expression system, we observed that final purification preparations have significantly more  $\alpha$ -subunit, skewing final protein subunit stoichiometry. Despite having a stoichiometry of 2 per molecule of RNAP, the  $\alpha$ -subunit appeared to be even more highly overexpressed compared to both the  $\beta$ - and  $\beta'$ -subunits. This is likely due to the tendency of both of these subunits to form inclusion bodies, reducing the amount of subunit available to form viable complex. The uneven prevalence of the  $\alpha$ -subunits in pre-chromatography sample lead to high occupancy of hexahistidine  $\alpha$ -subunit on NTA-Ni<sup>2+</sup> resin, leading to a significant amount of correctly assembled RNAP and excess  $\beta$ - and  $\beta'$ -subunits in the NTA-Ni<sup>2+</sup> flow through. This lead to the design of the 3<sup>rd</sup> generation expression system as well as a modified purification methodology.

In our 3<sup>rd</sup> generation expression system, the histidine-tag was extended from a hexahistidine to decahistidine tag and the *sigA* gene and *rpoZ*, encoding sigma factor A and the  $\omega$ -subunit were exchanged between their respective plasmids (see **Figure II-4A**). This was done in part to improve binding of the RNAP complex to the Ni<sup>2+</sup>-affinity resin and because the copy number of pRSF is significantly higher than pET and pAcYc. Having the  $\omega$ -subunit on a higher copy plasmid which leads to higher levels of overexpressed protein, this resulted in the observation of the  $\omega$ -subunit in soluble fractions of cell lysate (see **Figure II-4B**).

Purification methods for procurement of highly pure MTB RNAP holoenzyme were expanded to include cation-exchange, affinity, and anion-exchange chromatography steps (see **Appendix Figure II-A1**). In order to remove excess levels of decahistidine-tagged  $\alpha$ -subunit, the first chromatography step used was cation-exchange chromatography. RNAP holoenzyme is retained on negatively charged resin, while core

RNAP and excess  $\alpha$ -subunit flow through the resin (see **Appendix Figure II-A2**). Fractions containing MTB RNAP holoenzyme are then applied to a HisTrap HP, this allows for the removal of endogenous *E. coli* RNAP and is a further purification. RNAP yielded from this step of the purification is considerably pure (see **Appendix Figure II-A3**). The final polishing step in the purification is anion-exchange chromatography. This step, separates MTB RNAP into two peaks, one of which is significantly less active than the other (see **Appendix Figure II-A4**). This separation of two distinct complexes may indicate that the complex with lesser activity is likely misfolded. Additionally, this complex lacks  $\omega$ -subunit which has been reported to be critical for correct assembly of RNAP in mycobacteria as well as other bacterial RNAPs (see **Figure II-4B**).<sup>30</sup>

Initial characterization of the percentage of catalytically competent RNAP (those which are able to incorporate NTP into a nascent RNA in an artificial elongation complex) showed that a very small percentage of RNAPs were catalytically active. Reports in the literature which showed that a zinc-binding domain present in the  $\beta'$ -subunit played a critical role in *in vitro* reconstitution of RNAP complex prompted supplementation of growth media with zinc sulfate.<sup>22</sup> The drastic increase in catalytically competent RNAP complexes purified from cultures supplemented with excess zinc supports the observation that the zinc-binding domain likely plays a critical role in RNAP complex assembly (see **Figure II-5C**). Activity of the RMP<sup>R</sup> MTB RNAPs in the presence of 200 nM RMP show that we are not isolating chimeric MTB RNAP complexes containing *E. coli* RNAP (see **Appendix Figure II-A6**). Though chimeric RNAP formation did not seem to effect MTB RNAP, *E. coli* RNAPs which have a hexahistidine-tag on the C-terminus of the  $\beta'$ -subunit contained roughly ~25%  $\beta$ -subunit contamination (see **Appendix Figure II-A6**). This

would not likely effect IC<sub>50</sub> determinations for the *E. coli* RMP<sup>R</sup> mutants because complexes containing WT  $\beta$ -subunit would be fully inhibited at low concentrations of RMP.

Characterization of MTB and *E. coli* RMP<sup>R</sup> RNAPs susceptibility to challenge by RMP showed that point mutations in the RMP cleft are highly effective in reducing the potency of RMP (see **Figure II-6B**). Previous work which aimed to determine the IC<sub>50</sub> of RMP against both the WT and RMP<sup>R</sup> MTB and *E. coli* RNAPs used a single-stranded circular DNA template, whereas we are using a plasmid-based template which is more representative of physiological conditions.<sup>16</sup> A similar trend for the respective mutants in both *E. coli* and MTB can be observed. In both cases the H445Y (H526Y in *E. coli*) is essentially insensitive to RMP, even at 2 mM RMP (see **Figure II-6B**). The H526Y mutation (*E. coli* numbering) reshapes the RMP binding site and sterically blocks RMP binding to RNAP (see **Figure II-7C**), resulting in insensitivity to RMP. The inhibition by high concentrations of RMP that is seen for the *E. coli* H526Y RNAP is likely due to very low affinity RMP binding at a site distinct from the RRDR.

Using the WT RNAP•RMP complex structure as a template, several studies have previously modeled the RNAP structures containing RMP<sup>R</sup> mutations using molecular dynamics simulations or simple amino acid substitutions to postulate the molecular mechanisms of RMP<sup>R</sup> 16, 25-26, 31 The predictions about the D516V RNAP based on the *T. aquaticus* RNAP•RMP complex structure are consistent with the effects observed in the D516V RNAP•RMP complex in this work (see **Figure II-7B**). However, none of the previous work identified the RMP<sup>R</sup> mechanisms of the S531L and H526Y mutants which were revealed in our study.

Analysis of the structures of S531L RNAP with and without RMP revealed a bipartite mechanism resulting in the RMP<sup>R</sup> phenotype. The two aspects of the S531L RMP resistance are: 1) elimination of a key hydrogen bond between the S531 side chain and the phenol on C8 of the naphthalene ring; and 2) disordering of fork loop 2 that may result from a steric clash between RMP and the Leu side chain at position 531 (see **Figure II-7F**). The disorder of fork loop 2 disrupts RNAP•RMP contacts and also exposes hydrophobic segments of RMP to solvent, which is unfavorable for stable RNAP•RMP complex formation. In the absence of bound RMP, the S531L mutation did not exhibit any gross structural change in the RIF binding site. The significant structural changes of RRDR cluster I found in the H526Y mutant suggest that major modifications of the RMP ansa bridge backbone would be required for binding in the significantly altered RIF binding site of the H526Y mutant.

## Notes to Chapter II

Parts of this chapter were published in Molodstov et.al., *Molecular Microbiology*, **2017**, 103, 1034-1045. I would like to thank Dr. Vadim Molodstov and Dr. Katsuhiko Murakami for their work in elucidating the structures of the *E. coli* RMP<sup>R</sup> RNAP structures and for their critical analysis of the structural basis for each RMP<sup>R</sup> RNAP. I would also like to thank Dr. Nathan Scharf for his work in purifying and characterizing the *E. coli* RNAPs.

## References

1. Aristoff, P. A.; Garcia, G. A.; Kirchhoff, P. D.; Showalter, H. D., Rifamycins--obstacles and opportunities. *Tuberculosis (Edinb)* **2010**, *90* (2), 94-118.
2. Zumla, A.; Nahid, P.; Cole, S. T., Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov* **2013**, *12* (5), 388-404.
3. Poudel, A.; Nakajima, C.; Fukushima, Y.; Suzuki, H.; Pandey, B. D.; Maharjan, B.; Suzuki, Y., Molecular characterization of multidrug-resistant Mycobacterium tuberculosis isolated in Nepal. *Antimicrob Agents Chemother* **2012**, *56* (6), 2831-6.
4. Imperiale, B. R.; Zumarraga, M. J.; Weltman, G.; Zudiker, R.; Cataldi, A. A.; Morcillo, N. S., First evaluation in Argentina of the GenoType(R) MTBDRplus assay for multidrug-resistant Mycobacterium tuberculosis detection from clinical isolates and specimens. *Rev Argent Microbiol* **2012**, *44* (4), 283-9.
5. Rahmo, A.; Hamdar, Z.; Kasaa, I.; Dabboussi, F.; Hamze, M., Genotypic detection of rifampicin-resistant M. tuberculosis strains in Syrian and Lebanese patients. *J Infect Public Health* **2012**, *5* (6), 381-7.
6. Tang, K.; Sun, H.; Zhao, Y.; Guo, J.; Zhang, C.; Feng, Q.; He, Y.; Luo, M.; Li, Y.; Sun, Q., Characterization of rifampin-resistant isolates of Mycobacterium tuberculosis from Sichuan in China. *Tuberculosis (Edinb)* **2013**, *93* (1), 89-95.
7. Mani, N.; Dupuy, B.; Sonenshein, A. L., Isolation of RNA polymerase from Clostridium difficile and characterization of glutamate dehydrogenase and rRNA gene promoters in vitro and in vivo. *J Bacteriol* **2006**, *188* (1), 96-102.
8. China, A.; Nagaraja, V., Purification of RNA polymerase from mycobacteria for optimized promoter-polymerase interactions. *Protein Expr Purif* **2010**, *69* (2), 235-42.

9. Predich, M.; Doukhan, L.; Nair, G.; Smith, I., Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. *Mol Microbiol* **1995**, *15* (2), 355-66.
10. Helmann, J. D., Purification of *Bacillus subtilis* RNA polymerase and associated factors. *Methods Enzymol* **2003**, *370*, 10-24.
11. Borukhov, S.; Goldfarb, A., Recombinant *Escherichia coli* RNA polymerase: purification of individually overexpressed subunits and in vitro assembly. *Protein Expr Purif* **1993**, *4* (6), 503-11.
12. Xue, Y.; Hogan, B. P.; Erie, D. A., Purification and initial characterization of RNA polymerase from *Thermus thermophilus* strain HB8. *Biochemistry* **2000**, *39* (46), 14356-62.
13. Davis, E.; Chen, J.; Leon, K.; Darst, S. A.; Campbell, E. A., Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* **2015**, *43* (1), 433-45.
14. Harshey, R. M.; Ramakrishnan, T., Purification and properties of DNA-dependent RNA polymerase from *Mycobacterium tuberculosis* H37RV. *Biochim Biophys Acta* **1976**, *432* (1), 49-59.
15. Jacques, J. F.; Rodrigue, S.; Brzezinski, R.; Gaudreau, L., A recombinant *Mycobacterium tuberculosis* in vitro transcription system. *FEMS Microbiol Lett* **2006**, *255* (1), 140-7.
16. Gill, S. K.; Garcia, G. A., Rifamycin inhibition of WT and Rif-resistant *Mycobacterium tuberculosis* and *Escherichia coli* RNA polymerases in vitro. *Tuberculosis (Edinb)* **2011**, *91* (5), 361-9.
17. Banerjee, R.; Rudra, P.; Prajapati, R. K.; Sengupta, S.; Mukhopadhyay, J., Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. *Tuberculosis (Edinb)* **2014**, *94* (4), 397-404.
18. Temiakov, D.; Anikin, M.; McAllister, W. T., Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J Biol Chem* **2002**, *277* (49), 47035-43.



19. Scharf, N. T.; Molodtsov, V.; Kontos, A.; Murakami, K. S.; Garcia, G. A., Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase Discovered from High-Throughput Screening. *SLAS Discov* **2017**, *22* (3), 287-297.
20. Czyz, A.; Mooney, R. A.; Iaconi, A.; Landick, R., Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* **2014**, *5* (2), e00931.
21. Rong, M.; Durbin, R. K.; McAllister, W. T., Template strand switching by T7 RNA polymerase. *J Biol Chem* **1998**, *273* (17), 10253-60.
22. Markov, D.; Naryshkina, T.; Mustaev, A.; Severinov, K., A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes Dev* **1999**, *13* (18), 2439-48.
23. Arca, H. C.; Mosquera-Giraldo, L. I.; Pereira, J. M.; Sriranganathan, N.; Taylor, L. S.; Edgar, K. J., Rifampin Stability and Solution Concentration Enhancement Through Amorphous Solid Dispersion in Cellulose omega-Carboxyalkanoate Matrices. *J Pharm Sci* **2018**, *107* (1), 127-138.
24. Molodtsov, V.; Nawarathne, I. N.; Scharf, N. T.; Kirchhoff, P. D.; Showalter, H. D.; Garcia, G. A.; Murakami, K. S., X-ray crystal structures of the Escherichia coli RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem* **2013**, *56* (11), 4758-63.
25. Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A., Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **2001**, *104* (6), 901-12.
26. Artsimovitch, I.; Vassylyeva, M. N.; Svetlov, D.; Svetlov, V.; Perederina, A.; Igarashi, N.; Matsugaki, N.; Wakatsuki, S.; Tahirov, T. H.; Vassylyev, D. G., Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell* **2005**, *122* (3), 351-63.
27. Armstrong, J. A.; Hart, P. D., Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. *J Exp Med* **1975**, *142* (1), 1-16.

28. Wang, Y.; Li, C.; Khan, M. R.; Wang, Y.; Ruan, Y.; Zhao, B.; Zhang, B.; Ma, X.; Zhang, K.; Zhao, X.; Ye, G.; Guo, X.; Feng, G.; He, L.; Ma, G., An Engineered Rare Codon Device for Optimization of Metabolic Pathways. *Sci Rep* **2016**, *6*, 20608.
29. Kane, J. F., Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr Opin Biotechnol* **1995**, *6* (5), 494-500.
30. Mao, C.; Zhu, Y.; Lu, P.; Feng, L.; Chen, S.; Hu, Y., Association of omega with the C-terminal region of beta' subunit is essential for assembly of RNA polymerase in Mycobacterium tuberculosis. *J Bacteriol* **2018**.
31. Campbell, E. A.; Pavlova, O.; Zenkin, N.; Leon, F.; Irschik, H.; Jansen, R.; Severinov, K.; Darst, S. A., Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* **2005**, *24* (4), 674-82.

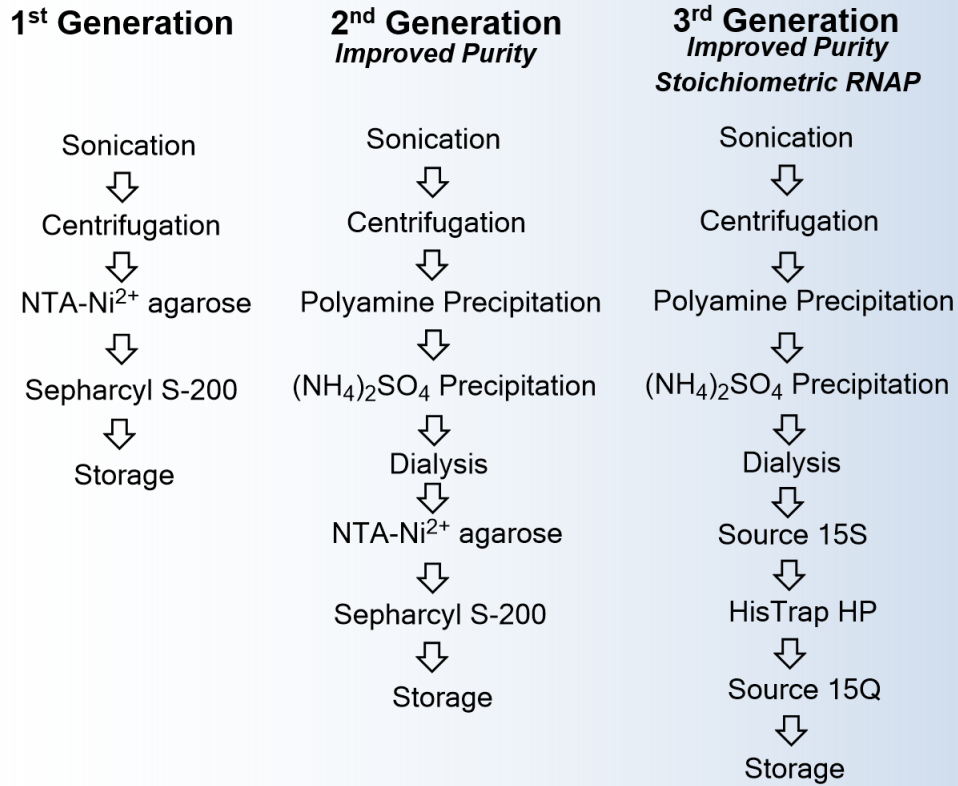
## Chapter II Appendix

Primer	Sequence	Purpose
<b>RpoC-MTBRP FOR</b>	GCCAACCTGGGAATCAATCTGTCCC	Amplify rpoC from pMTBRP for TOPO
<b>RpoC-MTBRP REV</b>	CCTGACGACGGATCGAACTGATCC	Amplify rpoC from pMTBRP for TOPO
<b>rpoA AMP FOR</b>	TATAC <b>CCATGG</b> TGATCTCACAGCGCCCCACCCTGTCC	Amplify rpoA from pMTBRP-NcoI
<b>rpoA AMP REV</b>	GGG <b>GAAGCTT</b> GGCTAAAGCTGTTTCGGTTTCGGCGTAGTCCTGCTCGTCCG	Amplify rpoA from pMTBRP-HindIII
<b>rpoB AMP FOR</b>	GAAC <b>CCCATGG</b> CAGATTCCCGCCAGAGCAAAACAGCCGC	Amplify rpoB from pMTBRP-NcoI
<b>rpoB AMP REV</b>	GAT <b>CGGATCC</b> TTAATGGTGATGGTGATGGTGATGGTGATGGTGCC CAAGATCCTCGACAC	Amplify rpoB from pMTBRP-BamHI
<b>rpoC AMP FOR</b>	GGGAGT <b>CATATG</b> CTCGACGTCAACTTCTTCGATGAACTCC	Amplify rpoC from pMTBRP-NdeI
<b>rpoC AMP REV</b>	CGTTATTTGAG <b>GATATC</b> CTAGCGGTAGTCGCTGTAGCCG	Amplify rpoC from pMTBRP-EcoRV
<b>rpoZ AMP FOR</b>	ATT <b>CATATG</b> AGTATCTCGCAGTCCGACGCG	Amplify rpoZ from pMTBRP-NdeI
<b>rpoZ AMP REV</b>	GAGT <b>GGTACC</b> GCTACTCGCCCTCGGTGTGC	Amplify rpoZ from pMTBRP-KpnI
<b>SigA AMP FOR</b>	GAAGGGGTGT <b>CCATGG</b> CAGCGACCAAAGCAAGCACGGCGACCGATGAGCC	Amplify SigA from gDNA for pRSF-NcoI
<b>SigA AMP REV</b>	GCCTCGGCGG <b>CAATTG</b> TCAGTCCAGGTAGTCGCGCAGGACCTGTGAGCGG	Amplify SigA from gDNA for pRSF-MfeI
<b>rpoA FOR NHis</b>	GAGAG <b>GATCC</b> TATGCTGATCTCACAGCGCCCCACCCTGTCC	Amplify rpoA for pAcYc-BamHI
<b>rpoB AMP 2 FOR</b>	ATAC <b>CATGG</b> TGTTGGCAGATTCCCGCCAGAGCAAAACAGC	Amplify rpoB without His-tag-NcoI
<b>rpoB AMP 2 REV</b>	CG <b>GGATCC</b> TTACGCAAGATCCTCGACACTTGCGGATTCGT	Amplify rpoB without His-tag-BamHI
<b>rpoA 1-229 AMP REV</b>	GATCT <b>CAAGCTT</b> CTAGGCCTCGACGTTGAGTTCCCGTGCC	Amplify truncated rpoA 1-229-HindIII
<b>rpoA AMP 3 REV</b>	GGCC <b>GGTACCTT</b> CTAAAGCTGTTTCGGTTTCGGCG	Amplify rpoA with no His tag for pAcYc-KpnI
<b>rpoA N10his FOR</b>	TATAC <b>CCATGG</b> GCAGCAGCCATCACCATCACCATCACCATCACCATC ACAGCCAGGATCCTATGCTGATCT	Amplify rpoA with N-terminal his10x - NcoI
<b>rpoZ NcoI AMP FOR</b>	CATTAC <b>CATGG</b> GATCTCGCAGTCCGACGCG	Amplify rpoZ with NcoI for pRSF
<b>rpoZ MfeI AMP REV</b>	GAGT <b>CAATTG</b> GCTACTCGCCCTCGGTGTGC	Amplify rpoZ with MfeI for pRSF
<b>SigA NdeI AMP FOR</b>	GAT <b>CCATATG</b> GCAGCGACCAAAGCAAGCACGG	Amplify SigA with NdeI for pAcYc
<b>SigA KpnI AMP REV</b>	GAT <b>CGGTACCT</b> CAGTCCAGGTAGTCGCGCAGG	Amplify SigA with KpnI for pAcYc

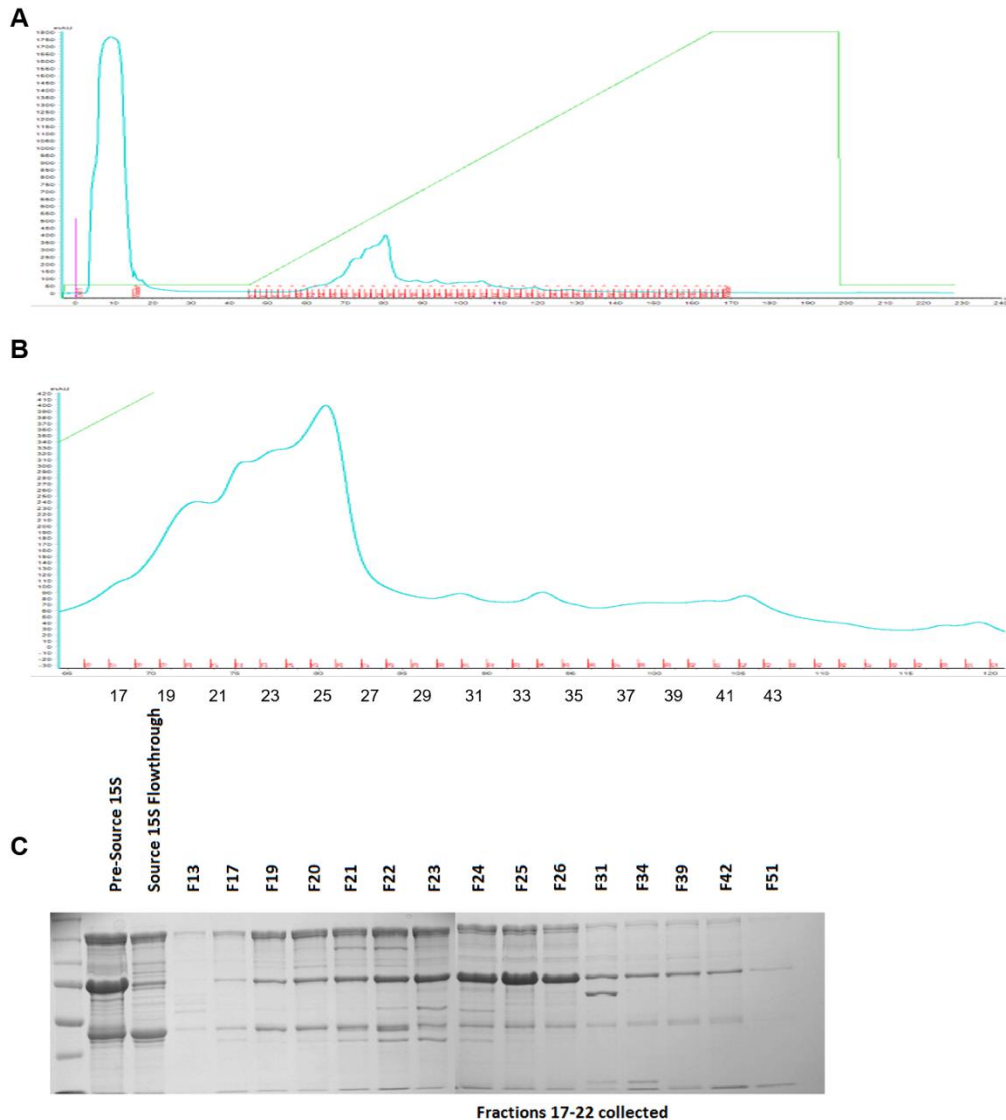
- Restriction sites are in bold

**Appendix Table II-A1: Primers used in Chapter II**

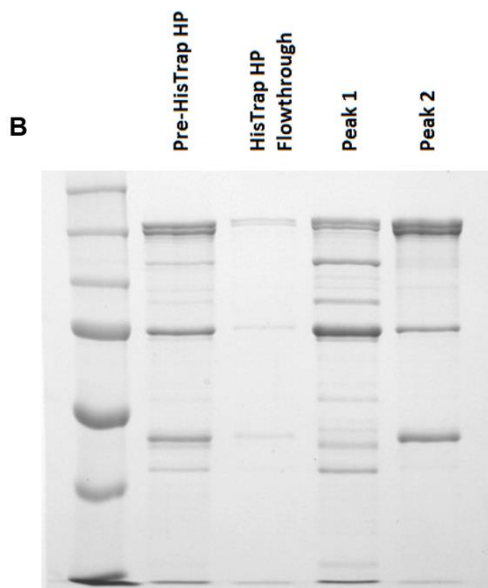
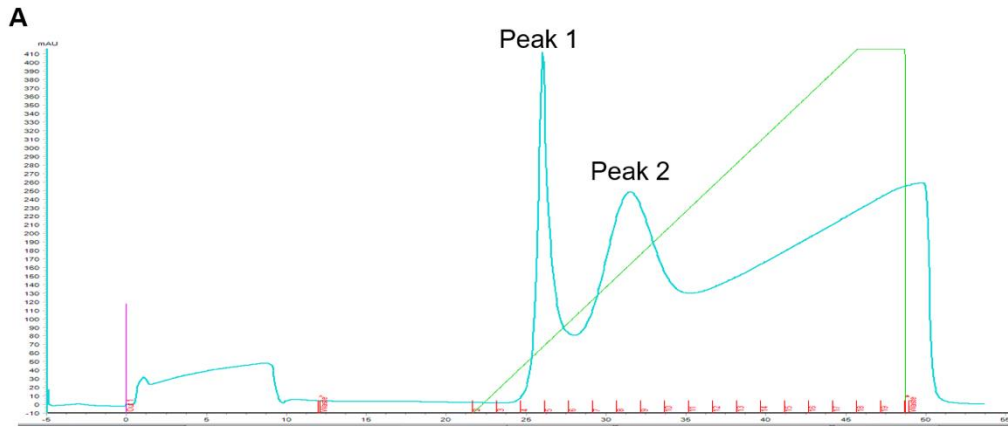
## Purification of MTB RNAP



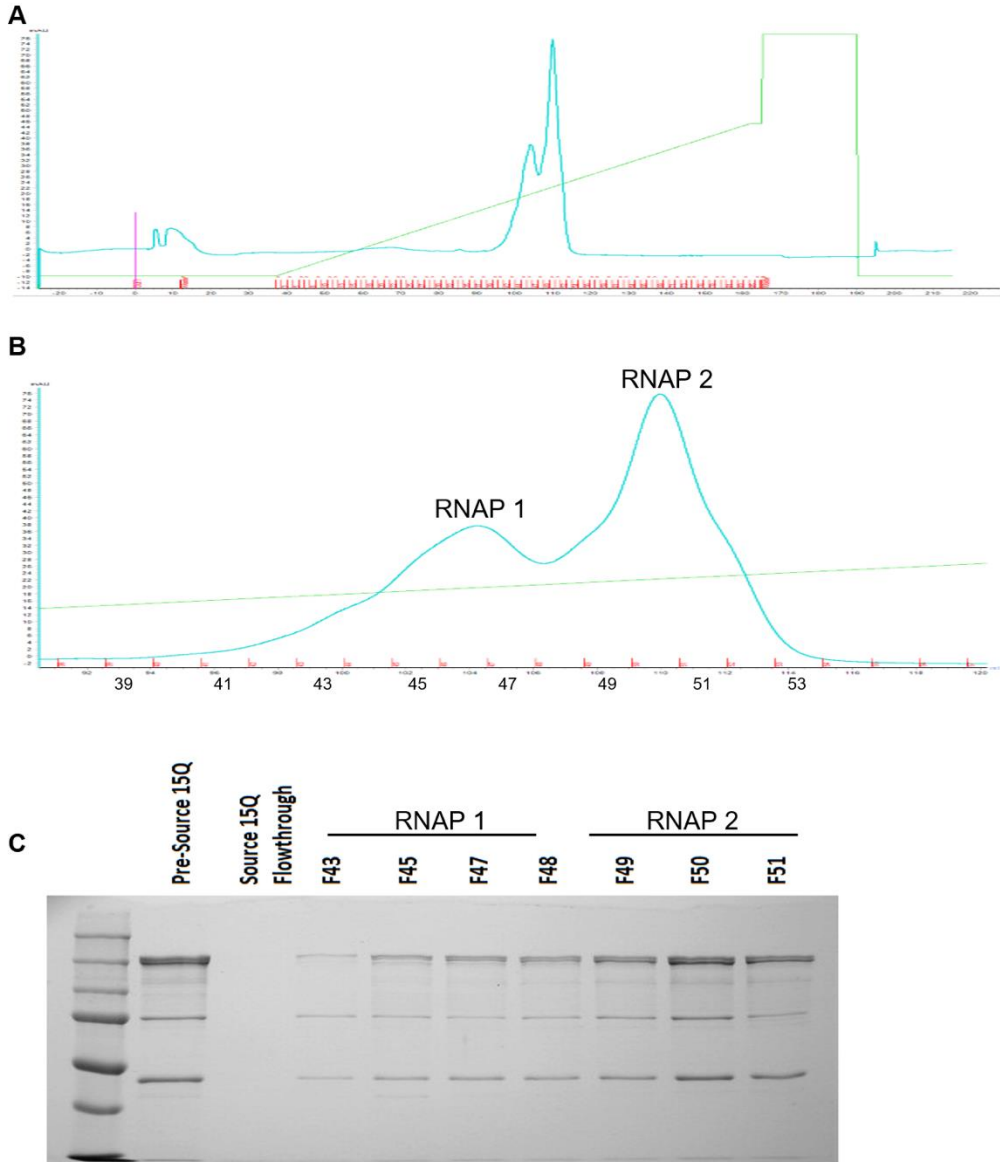
Appendix Figure II-A1: Schematic showing the evolution of the purification of MTB RNAP



**Appendix Figure II-A2: Purification of MTB RNAP with cation-exchange chromatography.** **A)** Representative chromatogram of MTB RNAP purification using a 6mL Source 15S column (GE Healthcare). **B)** Close up of peaks eluted from A. **C)** SDS-PAGE MTB RNAP purification using the Source 15S column including pre-injection, Source 15S flow through, and eluted fraction samples.

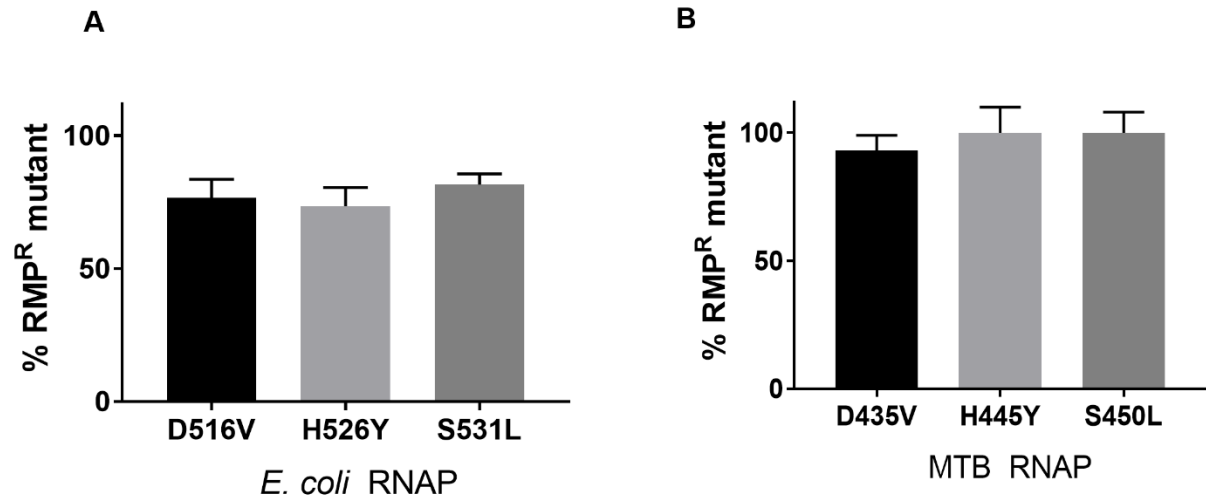


**Appendix Figure II-A3: Purification of MTB RNAP with immobilized-metal affinity chromatography. A)** Representative chromatogram of MTB Purification using a 1mL HisTrap HP column (GE Healthcare). **B)** SDS-PAGE of MTB RNAP purification using the HisTrap HP column including pre-injection, HisTrap flow through, and eluted peak samples.



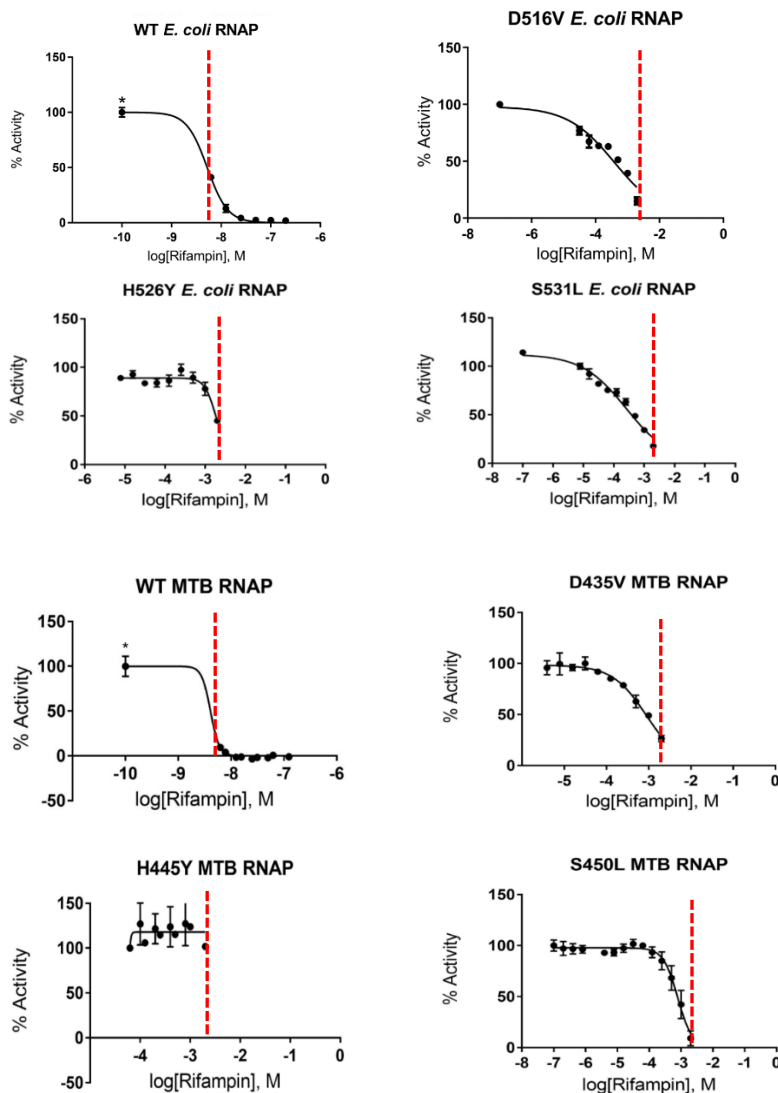
**Appendix Figure II-A4: Purification of MTB RNAP with anion-exchange chromatography.**

**A)** Representative chromatogram of MTB RNAP purification using a 6mL Source 15Q column (GE Healthcare). **B)** Close up of peaks eluted from A. **C)** SDS-PAGE MTB RNAP purification using the Source 15Q column including pre-injection, Source 15Q flow through, and eluted fraction samples.



**Appendix Figure II-A5: Determination of percent of contaminating endogenous WT *E. coli* RNAP in the RNAP purifications. A) Contamination in *E. coli* RMP<sup>R</sup> and B) MTB RMP<sup>R</sup> RNAP preparations. Essentially no endogenous RNAP contamination was found in the MTB preparations.**





**Appendix Figure II-A6: Rifampin IC<sub>50</sub> curves for WT and RMP<sup>R</sup> *E. coli* and MTB RNAPs. (A) *E. coli* RNAPs (WT and RIFR mutants) and (B) MTB RNAP (WT and RIFR mutants). All of the average data points (•) were plotted. The average data were fit by nonlinear regression, with upper and lower limits set to 100 and 0 respectively, to determine their IC<sub>50</sub> values. In the WT RNAP plots, the values at log = -10, annotated with a single asterisk, are the DMSO controls. The dashed line indicates the limits of detection, either from enzyme concentration limitations needed for signal or from RMP solubility.**

## CHAPTER III

### ***In vitro* Characterization of Rifampin Resistant *M. tuberculosis* RNA polymerases**

Development of resistance to antibiotics usually comes at some cost to the organism's fitness<sup>1-2</sup>. A significant fitness defect, coupled with prolonged exposure to antibiotics, creates an environment where selective pressure to compensate for fitness defects can lead to the development of secondary mutations. There have been extensive studies correlating the prevalence of RMP<sup>R</sup> mutations in MTB clinical isolates with the fitness costs of rifamycin-resistance<sup>3-6</sup>. Most of these studies employed either competition assays or growth rate measurements to determine if a resistance mutation causes a fitness defect<sup>7</sup>. Though many of the studies have tried to identify correlations between fitness and genotype, differences in genetic background and artificial growth conditions may convolute how fitness is presented. Additionally, pleiotropic effects from drug resistance mutations can further complicate interpretation of direct effects from a particular mutation. In contrast, the intrinsic changes in enzyme function *in vitro* can be interpreted as a direct effect of a particular mutation.

Laboratory-evolved strains carrying the same RMP<sup>R</sup> point mutations have been observed to exhibit varying fitness defects. Interestingly, strains carrying the  $\beta$ S450L mutation tend to exhibit the least fitness defect when compared to other RMP<sup>R</sup> mutants in competition studies. The minimal observed fitness defect for this particular mutant was used as rationale for explaining its prevalence in clinical isolates<sup>3</sup>. However, a number of clinical isolates containing the  $\beta$ S450L mutation were observed to have varying fitness

defects despite sharing the same RMP resistance allele. This led to the discovery of secondary mutations, predominantly at the interface of the  $\alpha$ - and  $\beta'$ -subunits of RNAP, a majority of which are found in the double-psi beta-barrel (DPBB) domain of the  $\beta'$ -subunit, the distal end of which contributes to the catalytic site of RNAP and includes the magnesium-coordinating motif, DFDGD (see **Figure I-5**). We refer to these secondary mutations as “compensatory mutations” for their ability to ameliorate/compensate for the fitness defects of the primary RMP<sup>R</sup> mutants. The presence of these compensatory mutations is overwhelmingly, but not exclusively associated with the  $\beta$ S450L RMP<sup>R</sup> mutation (75%), approximately 9% of the compensatory mutations were observed in the *rpoA* or  $\alpha$ -subunit <sup>7-8</sup>. The  $\beta$ V483G and  $\beta$ F452L mutations are the most prevalent mutations associated with the  $\beta$ S450L mutant and have been observed in all of the studies mentioned above, suggesting that they arose by convergent evolution and are physiologically relevant to the organism <sup>3</sup>.

There is an incomplete understanding regarding the disproportionate representation of particular RMP<sup>R</sup> mutations in clinical isolates despite the similar efficacies of these mutations to decrease the susceptibility of RNAP to RMP <sup>9</sup>. Our study provides a molecular foundation for the enrichment of the  $\beta$ S450L mutant in clinical isolates and elucidates fitness defects associated with this mutant which likely drive the pressure to develop compensatory mutations. Three MTB trans-acting transcription factors, CarD, NusA, and NusG, were also studied to determine the extent of alteration of the intrinsic properties of each RMP<sup>R</sup> mutant in physiologically relevant contexts. Several studies of RMP<sup>R</sup> *E. coli* have shown that mutations in the RRDR have differential effects on transcription initiation, elongation, and termination <sup>10-13</sup>; however, these

mutants have usually been selected from phenotypic screens of laboratory evolved RMP<sup>R</sup> mutants and differ from those observed in clinical settings. Until now these effects have not been biochemically characterized in clinically relevant MTB RMP<sup>R</sup> RNAPs. These studies provide insight into aspects of transcription which are critical to the MTB viability.

## Materials and Methods

### *Plasmid and DNA Manipulation*

Plasmid pMt-rrnA3 contains the MTB *rrnA3* promoter region -66 to +351. DNA was amplified from H37Rv genomic DNA (+1468189 to +1468605) and subcloned into pTZ18U using PciI and EcoRI restriction sites using primers Mt *rrnA3* -66 AMP FOR and Mt *rrnA3* +351 AMP REV (see **Appendix Figure III-A1** and **Appendix Table III-A1** for all primers). pMt-rrnA3-synB was produced by insertion of the synB terminator derived from synthetic oligonucleotides (IDT Technologies) with flanking XbaI restriction sites (see **Appendix Table III-A1** for all plasmids). pMt-rrnA3 derivative, pEC19, was produced by mutagenesis of both adenosines A10 and A11 (relative to the *rrnA3* TSS) to thymine using *rrnA3* EC19 FOR and *rrnA3* EC19 REV. pEC26-Tuf was generated by digestion of pMt-rrnA3 with PciI and MscI to remove the MTB *rrnA3* promoter. The *E. coli* consensus promoter which lacks any cytosine until the +26 position, relative to the transcription start site, was subcloned using synthetic oligonucleotides. The MTB *tuf* and *metK* terminators were subcloned into pEC26 using synthetic oligonucleotides containing the terminator sequences with flanking XhoI restriction sites. Plasmids used for promoter studies were purified using a Qiagen Miniprep kit. PCR amplified DNA fragments were generated from pEC19 (elongation studies) and pEC26-Tuf or pEC26-*metK* (termination studies) using primers EC19 AMP FOR and EC AMP REV for the elongation studies template and EC26

AMP FOR and EC AMP REV for the termination studies. PCR products were purified on a 2% agarose gel and extracted with Qiagen gel extraction kit. Plasmid DNA and PCR fragments were then further purified by phenol-chloroform extraction followed by ethanol precipitation. DNA was stored in water at -20°C.

Expression constructs for MTB CarD (Rv3583c), NusA (Rv2841c), and NusG (Rv0639) were produced similarly. Each gene was amplified from H37Rv genomic DNA using primers in **Appendix Table III-A1**. All genes were subcloned into a modified pET19b construct containing a N-terminal decahistidine tag followed by the Precision Protease recognition cleavage site LEVLFQ/GP (pET19bpps). CarD and NusA were subcloned into the pET19bpps vector using NdeI and BamHI restriction sites. NusG was subcloned into pET19bpps using the NdeI and XhoI restriction sites (see **Appendix Table III-A2**).

#### *Protein Purification*

MTB RNAP was prepared as previously reported<sup>9</sup>. CarD was prepared as follows. BL21(DE3) *E. coli* containing pET19bpps-CarD were grown to an OD<sub>600</sub> of 0.8. Protein expression was induced by addition of 1mM IPTG. Cells were incubated at 28°C for 4 hours prior to harvesting via centrifugation. Cells were lysed via sonication in 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5% glycerol, 5 mM β-ME (β-mercaptoethanol), 1 mM PMSF and 1X Roche cOmplete ULTRA protease cocktail), 200U DNaseI, and 1 mg/mL lysozyme. Clarified lysate was passed through a 0.22 μm filter and applied to a HisTrap HP column (GE Healthcare). CarD was eluted over linear gradient to 500 mM imidazole. Fractions containing decahistidine-tagged CarD were collected and Precision Protease (GE Healthcare) was added. CarD was dialyzed overnight at 4°C in 10 mM Tris-HCl (pH

8.0), 500 mM NaCl, 0.1 mM EDTA, 5% glycerol, 5 mM  $\beta$ -ME. CarD was then passed through 1 mL of GST-agarose to remove Precision Protease and concentrated prior to application to a HiPrep Sephacryl S-200 HR column (GE Healthcare) run in 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.1 mM EDTA, 5% glycerol, and 1 mM DTT. Purified CarD was stored at  $-80^{\circ}\text{C}$  (see **Appendix Figure III-A2**).

NusA and NusG were prepared identically. *E. coli* BL21 (DE3) cells containing either pET19bpps-NusA or pET19bpps-NusG were grown to an  $\text{OD}_{600}$  of 0.4 and induced by addition of 0.5 mM IPTG. Cells were incubated overnight at  $16^{\circ}\text{C}$ . Cells were lysed via sonication in 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5mM imidazole, 5% glycerol, 5 mM  $\beta$ -ME, 0.1% Triton X-100, 1 mM PMSF and 1X Roche cOmplete ULTRA protease cocktail), 200U DNaseI, and 1mg/mL lysozyme buffer. Clarified lysate was passed through a 0.22 $\mu\text{m}$  filter and then applied to a HisTrap HP column and proteins were eluted by a linear gradient to 500mM imidazole. Fractions containing either Nus protein were collected and Precision Protease was added. Proteins were dialyzed into 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mM  $\beta$ -ME, 5% glycerol, 0.1 mM EDTA and passed through 1 mL of GST-agarose before application to a 6 mL Source 15Q column equilibrated with TGEB (10 mM Tris-HCl (pH 8.0), 5% glycerol, 0.1 mM EDTA, 5 mM  $\beta$ -mercaptoethanol) with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl. Fractions containing Nus proteins were collected and dialyzed into 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol. Proteins were stored at  $-80^{\circ}\text{C}$  (see **Appendix Figure III-A2**).

### *Determination Open-promoter Complex Half-life*

Open-promoter complex half-life was determined by single-round transcription. Supercoiled plasmid template containing -66 to +350 of the transcription start site of the MTB *rrnA3* promoter followed by the *synB* terminator was used for this assay. MTB RNAP (30 nM) was pre-incubated with pMt-*rrnA3*-*SynB* (10 ng/ $\mu$ L) for 30 minutes at 37°C to form the open-complex in 40mM Tris-HCl (pH 8.0), 10 mM MgCl<sub>2</sub>, 30 mM KCl, 1 mM DTT, 0.1 mM EDTA, and 25  $\mu$ g/mL BSA. When present, CarD was added to a final concentration of 2  $\mu$ M and incubated for an additional 5 minutes. Heparin was then added to 50  $\mu$ g/mL and aliquots taken at various time points were added to the NTP mixture (1 mM GTP, 200  $\mu$ M ATP, 200  $\mu$ M UTP, 10  $\mu$ M CTP, 0.15  $\mu$ Ci/uL [ $\alpha$ <sup>32</sup>P]CTP final concentration). Reactions were allowed to continue for 15 minutes before addition of 2X RNA loading buffer (95% formamide, 0.02% SDS, 0.01% bromophenol blue, 0.005% xylene cyanol, 0.5 mM EDTA). Reactions were resolved via 6% acrylamide, 8 M urea PAGE. Gels were imaged and bands quantified by phosphorimaging (Molecular Dynamics, ImageQuant software). The fraction of complex at each time point was normalized to the zero time point and fitted to a linear plot. The half-life was determined by calculating the time at which half of the remaining complex was present. All reactions were conducted in triplicate with the errors representing the standard deviation of the slope of the line.

### *Determination of elongation rate*

Elongation rate was determined by a single-round transcription assay on a linear template. MTB RNAP (120 nM) was pre-incubated with linear DNA (30 nM), EC19, in 40 mM Tris-HCl (pH 8.0), 10 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 0.1 mM EDTA, and 25 µg/mL BSA for 15 minutes at 37°C. Stalled-elongation complex was formed by incubating 10 µM GTP, 10 µM UTP, 2 µM CTP, and 0.3 µCi/µL [ $\alpha^{32}$ P] CTP for 10 minutes. Transcription was resumed by adjusting the all NTPs to 100 µM and addition of 50 µg/mL heparin. Aliquots were removed at the time points denoted in the figures and stopped by addition of 2X RNA loading buffer. Reactions were resolved via 8% acrylamide, 8 M urea PAGE and processed as mentioned above. Elongation rates were determined by quantifying the density of run-off product over time which was fit to a standard Boltzmann sigmoidal curve. Average elongation rates were determined as the point at which 50% of runoff product is formed and is presented in nucleotides per second. All experiments were conducted in triplicate.

### *Determination of termination efficiency*

Termination efficiencies were determined by single-round transcription assays on linear templates. MTB terminators *tuf* and *metK* were identified by genomic profiling of MTB intrinsic terminators<sup>14</sup>. MTB RNAP (50 nM) was pre-incubated with linear DNA (25 nM), EC26-*metK* or EC26-*Tuf*, in 40 mM Tris-HCl (pH 8.0), 10 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 0.1 mM EDTA, and 25 µg/mL BSA for 15 minutes at 37°C. Stalled elongation complexes were formed by incubating 10 µM GTP, 10 µM ATP, 2 µM UTP, and 0.1 µCi/µL [ $\alpha^{32}$ P]UTP for 10 minutes. When used, NusA and NusG were added to 1 µM and allowed to incubate for 10 minutes. Transcription was restarted by addition of 50 µg/mL heparin



and adjusting each NTP concentration to 200  $\mu$ M. Reactions were allowed to continue for 30 minutes before 2X RNA loading buffer was added. Reactions were resolved on 6% acrylamide, 8M urea PAGE and processed as mentioned above. Termination efficiencies were calculated by dividing the density of terminated products by the sum of the terminated and runoff products in each lane.

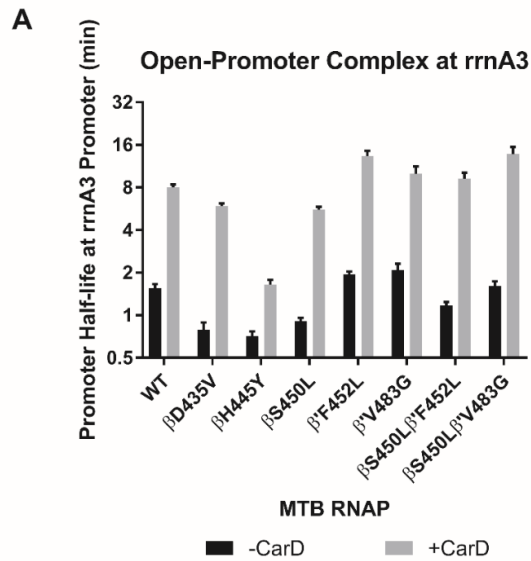
### *Primer extension Assays*

Determination of the concentration of active RNAP and percent of RNA hydrolysis was conducted using a DNA•RNA scaffold that mimics the transcription elongation complex.<sup>15</sup> First, an 8 nucleotide RNA (IDT technologies) was labeled at the 5' end with [ $\gamma$ P<sup>32</sup>] GTP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs). The T4 polynucleotide kinase was then heat inactivated. The DNA•RNA heteroduplex scaffolds were prepared by mixing equimolar amounts of template DNA26 (5'-CTGAAATCGACATCGCCGCTCAACAT-3') and 5'-labeled RNA8 (5'-GCGGCGAU-3'), heating to 85 °C for 10 min, followed by slow cooling to room temperature. The elongation complex was formed by mixing equimolar RNAP with the DNA•RNA heteroduplex scaffold (500 nM) in reaction buffer at 28 °C for 10 min. The reaction consisted of 40 mM Tris-HCl (pH 7.9), 40 mM potassium glutamate, 0.1 mM EDTA, 10 mM MgCl<sub>2</sub>, 25  $\mu$ g/mL BSA, 1 mM DTT; both reactions contained 15 mU of inorganic pyrophosphatase (New England Biolabs). The RNA extension reaction was conducted for 10 min after adding the next incoming GTP (400  $\mu$ M final concentration) to the elongation complex. RNA extension was stopped by adding an equal volume of 2x RNA loading buffer. RNAs were resolved by gel-electrophoresis in 20% acrylamide, 8 M urea PAGE gel and processed as above.

## Results

*RMP<sup>R</sup> mutant RNAPs form unstable open-promoter complexes which are stabilized by compensatory mutations in the  $\beta'$ -subunit.*

Single-round transcription assays performed on supercoiled plasmid DNA template containing the MTB ribosomal RNA promoter (*rrnA3*) upstream of the *synB* terminator <sup>16</sup> were used to determine open-complex half-lives in the presence and absence of CarD (see **Figure III-1**, **Appendix Figure III-A3**), a transcriptional regulator which is known to stabilize the *rrnA3* open-promoter complex. The open-promoter complex half-life ( $t_{1/2}$  = 1.5 min) for the wild-type (WT) MTB RNAP determined in the absence of CarD under these conditions is consistent with previous reports in the literature ( $t_{1/2}$  = 2 min) <sup>17</sup>. All single RMP<sup>R</sup> mutations reduced the stability of the RNAP open-promoter complexes, having half-lives ~50% of that for the WT. The compensatory mutations,  $\beta'$ F452L and  $\beta'$ V483G, provide stabilizing effects (20-30% longer half-life) on the WT open-promoter complex. These compensatory mutations also mitigated the defect in open-promoter complex stability for the  $\beta$ S450L mutant. Open-promoter complex half-lives were not evaluated for the  $\beta$ D435V and  $\beta$ 445Y mutants with compensatory mutations because these mutants are not observed in clinical isolates. Similar to in vivo observations, the  $\beta'$ F452L mutation partially recovered this defect (from 50% to 75% of WT) and the  $\beta'$ V483G mutation fully recovered the stability defect, bringing the open-complex half-life for the  $\beta$ S450L $\beta'$ V483G double mutant essentially equal to that of the WT <sup>18</sup>.



**B**

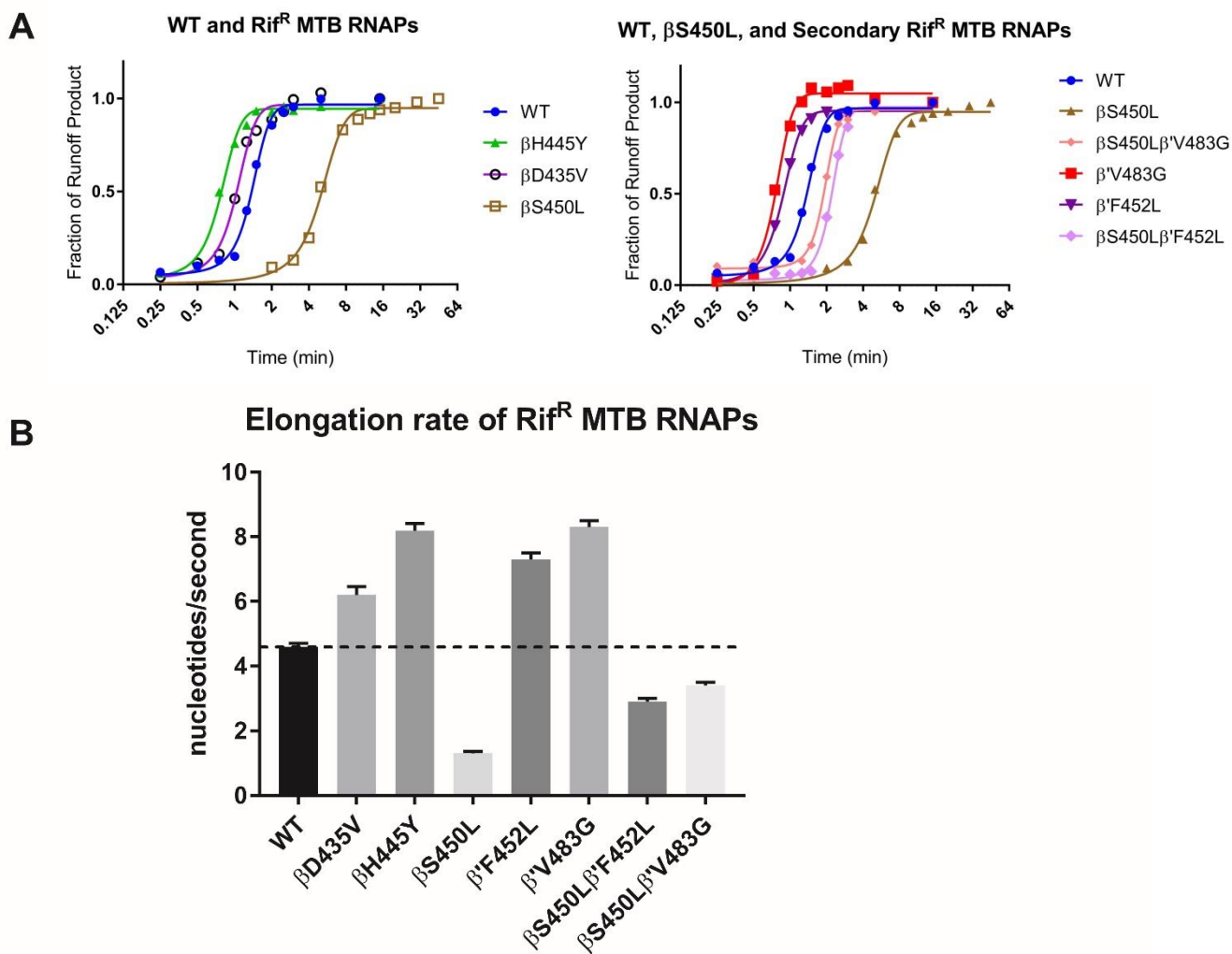
RNAP Mutant	$t_{1/2}$ (min)		Relative $t_{1/2}$ +CarD/-CarD
	-CarD	+CarD	
WT	1.6 ± 0.1	8.0 ± 0.4	5.0
βD435V	0.8 ± 0.1	5.9 ± 0.3	7.4
βH445Y	0.7 ± 0.1	1.7 ± 0.1	2.4
βS450L	0.9 ± 0.1	5.6 ± 0.3	6.2
β'F452L	1.9 ± 0.1	13.4 ± 1.2	7.2
β'V483G	2.1 ± 0.2	10.0 ± 1.3	4.8
βS450Lβ'F452L	1.2 ± 0.1	9.3 ± 0.9	7.8
βS450Lβ'V483G	1.6 ± 0.1	13.8 ± 1.7	8.6

**Figure III-1: Open-promoter complex studies with MTB RNAP in the presence and absence of CarD. A)** Bar graph of open-promoter complex half-lives at the MTB *rrnA3* promoter in the presence and absence of transcription factor CarD. (Note: Promoter half-life times are in Log2 scale). **B)** Table of open-promoter complex half-lives of MTB RNAP variants in presence and absence of CarD.

In the presence of CarD the WT open-promoter complex was stabilized roughly 5-fold (see **Figure III-1B** and **Appendix Figure III-A3**). CarD also stabilized all RNAP mutant half-lives 5- to 8-fold, with the exception of the  $\beta$ H445Y mutant which exhibited only a 2.4-fold increase in half-life. A similar trend in open-complex half-life across the RNAP variants is observed in the presence of CarD. Whether CarD is present or not, all RNAPs containing a compensatory mutation in the  $\beta'$ -subunit have a half-life greater than that observed for WT.

*Compensatory mutations in the  $\beta'$ -subunit mitigate the defect in elongation rate for the  $\beta$ S450L RMP<sup>R</sup> mutant.*

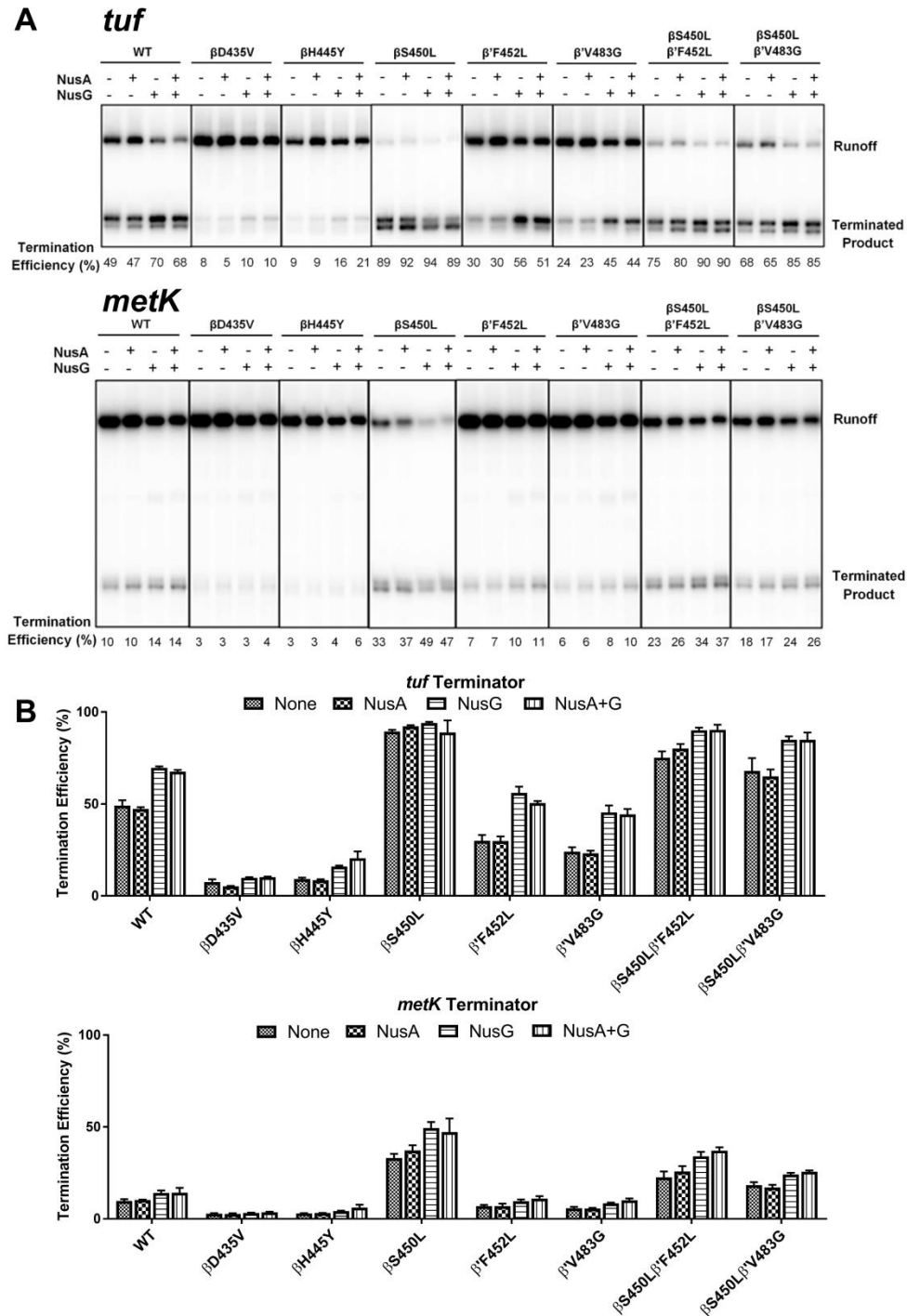
Elongation rates for WT and each of the MTB RNAP mutants were determined using single-round transcription assays on a linear template containing the *rmA3* promoter (see **Figure III-2**, **Appendix Figure III-A4**). Run-off transcripts from this template produce a 398 nucleotide RNA. The template was modified by substitution of A10 and A11 each with T to allow for formation of a stalled elongation complex at the +19 position (EC19).  $\beta$ D435V and  $\beta$ H445Y RMP<sup>R</sup> RNAPs have greater elongation rates relative to WT, whereas the  $\beta$ S450L mutant has a significantly slower elongation rate ( ). Again, the presence of a compensatory mutation in the  $\beta'$ -subunit was able to substantially mitigate this defect (see **Figure III-2B**). Each compensatory mutation on its own resulted in an increase in elongation rate relative to WT. It is noteworthy that there is a lack of significant pause sites along the span on the transcribed sequence downstream of the stalled +19 site. This allows the observed elongation rates to be attributed to processivity (phosphodiester bond formation) and not confounded by time spent at stalled positions along the template (see **Appendix Figure III-A4**).



**Figure III-2: Elongation Rate of WT and RMP<sup>R</sup> MTB RNAPs.** **A)** Plots used to determine the elongation rate for each of the RNAPs in this study. Density of runoff transcript was plotted against time. The time at which 50% of elongating RNAPs reach runoff length was divided by the length of the transcribed region to determine the average elongation rate. **B)** Summary of MTB RNAP mutants' elongation rates determined from the curves in A and B.

*The  $\beta$ S450L mutation improves termination efficiency, while the  $\beta$ D435V and  $\beta$ H445Y mutations cause defective intrinsic termination at the *tuf* and *metK* terminators.*

Termination efficiency was determined by single-round transcription on a linear template containing the *tuf* or the *metK* terminator downstream of a strong *E. coli* consensus promoter<sup>16</sup>. RMP<sup>R</sup> mutant RNAPs had varying termination efficiencies (see **Figure III-3**) which inversely correlated with elongation rates (see **Figure III-6B**). For both the *metK* and *tuf* terminators, there are two adjacent termination sites (consistent with the reported termination pattern for the *tuf* terminator<sup>16</sup>), the bands for both of which were included in the calculation of termination efficiency (see **Figure III-3A**). The  $\beta$ D425V and  $\beta$ H445Y RMP<sup>R</sup> were less able than WT to terminate at either terminator. RNAPs containing the  $\beta$ S450L mutation overall were clearly the most efficient at intrinsic termination. The presence of either the  $\beta$ 'F452L or  $\beta$ 'V483G mutation decreased the termination efficiency of the WT RNAPs. However, both double mutants ( $\beta$ S450L/ $\beta$ 'F452L and  $\beta$ S450L/ $\beta$ 'V483G) retained higher termination efficiencies than WT, albeit lower than the  $\beta$ S450L single mutant.



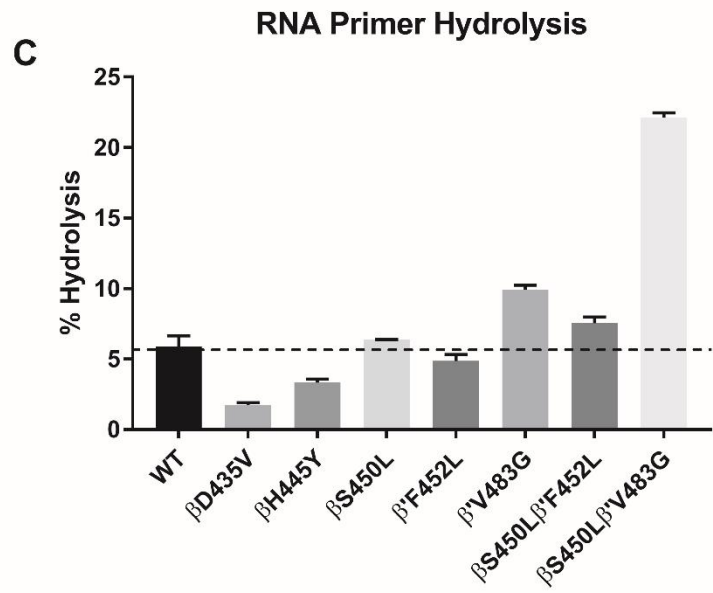
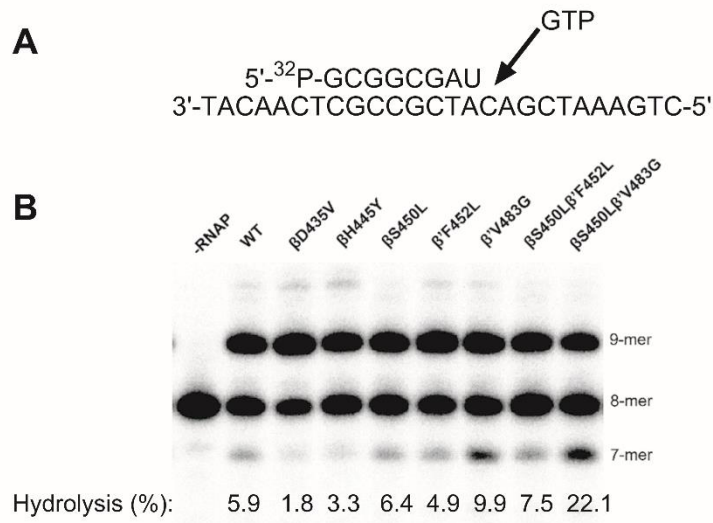
**Figure III-3: Termination studies at the *tuf* and *metK* terminators with all MTB RNAPs included in this study in the presence and absence of NusA and NusG. A) Phosphorimage gels used to calculate termination efficiency. B) Summary of termination efficiencies of WT and RMP<sup>R</sup> MTB RNAPs.**

Trans-acting termination factors, NusA and NusG (Rv0639), were also included in termination assays. NusA stimulates termination<sup>19</sup>. Rv0639 was annotated as the NusG homolog in MTB; however, characterization of the protein indicates it diverges from the canonical function observed for NusG in *E. coli* by stimulating termination in MTB<sup>16, 20</sup>. MTB rv0639/NusG had varying effects on termination depending on the terminator and RNAP mutant. At the *tuf* terminator, the RNAPs containing no RMP<sup>R</sup> mutations exhibited the strongest increases in termination efficiencies upon addition of NusG relative to other RNAPs. This was not the case at the *metK* terminator. At the *metK* terminator, RNAPs containing the  $\beta$ S450L mutation had the greatest, albeit modest relative to the *tuf* terminator, increases in termination efficiencies relative to no NusG, although this increase was less than that for WT with the *tuf* terminator. The  $\beta$ S450L mutant was highly efficient at termination in the absence of Nus factors. NusA, which has been shown to increase termination efficiency *in vitro* for *E. coli*<sup>21-22</sup>, had essentially no effect on termination efficiencies. NusG stimulates pausing along the template in the elongation assay (see **Appendix Figure III-A5**).

*Compensatory mutations in the DPBB enhance RNA primer hydrolysis.*

Single nucleotide hydrolysis of, and incorporation into an RNA8:DNA26 heteroduplex scaffold were monitored by denaturing PAGE (see **Figure III-4A**)<sup>15</sup>. We observed varying degrees (all much smaller than incorporation) of hydrolysis of the 3' UMP of the RNA8 oligonucleotide to a 7-mer product (see **Figure III-4B-C**). This likely happened during the 10-minute preincubation prior to the addition of GTP. RNAP double mutants with the  $\beta$ S450L mutation and a compensatory mutation in the  $\beta'$ -subunit had elevated levels of primer hydrolysis to the 7-mer, with the  $\beta'$ V483G mutant





**Figure III-4: Percent hydrolysis of WT and RMP<sup>R</sup> MTB RNAPs. A)** RNA•DNA scaffold used in this study. **B)** Radiograph showing primer extension and hydrolysis of RNA8 primer. **C)** Percent hydrolysis for each RNAP in this study

(both with and without the  $\beta$ S450L mutation) having the greatest propensity for hydrolysis. The  $\beta$ D435V and  $\beta$ H445Y RMP<sup>R</sup> mutants displayed lower levels of hydrolysis, while the  $\beta$ S450L mutant displayed a similar proclivity for hydrolysis to that of the WT (see **Figure III-6C**). Upon addition of GTP, the RNA8 is extended to the 9-mer. (Any hydrolysis of the 9-mer back to the 8-mer would be immediately corrected by addition of another G due to the excess of GTP (800-fold) present.) The percentage of active (e.g., catalytically competent) complex in a given preparation of RNAP was determined as the sum of the extension to the 9-mer and hydrolysis to the 7-mer RNA. MTB RNAPs used in this study had activities ranging from 45-77% (see **Figure III-6C, Appendix Table III-A3**).

### Discussion

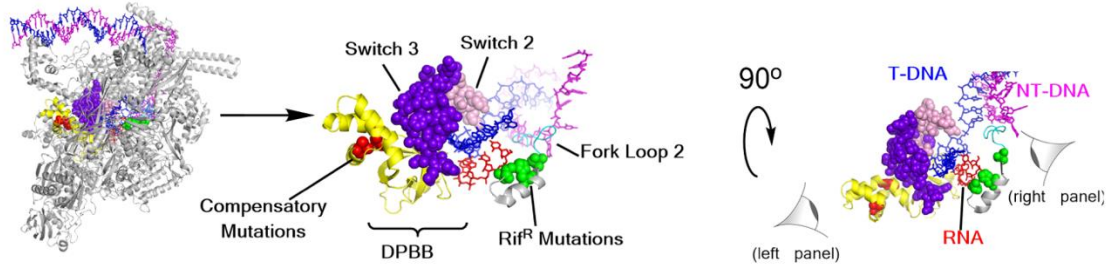
Rifampin resistant (RMP<sup>R</sup>) mutations in the rifamycin-resistance determining region (RRDR) have been observed to affect open-promoter complex (RPo) stability in the *E. coli* RNAP<sup>11</sup>. We find that RMP<sup>R</sup> mutations in the  $\beta$ -subunit of MTB RNAP ( $\beta$ D435V,  $\beta$ H445Y, and  $\beta$ S450L) all share similar reductions in RPo complex stabilities relative to WT RNAP. CarD, a transcription factor that significantly (5-8-fold) increases RPo stability for all MTB RNAPs studied, had minimal effects on the trends observed for the WT and RMP<sup>R</sup> mutants in the absence of CarD, with the exception of the  $\beta$ H445Y mutant which had a lower fold change in promoter half-life (see **Figure III-1**). We have previously shown that the  $\beta$ H445Y MTB RNAP exhibits a significantly altered RMP binding pocket conformation<sup>9</sup> and the greatest reduction in RPo stability relative to WT. This may reduce the ability of CarD to fully stabilize the RPo complex for this mutant.

The RMP<sup>R</sup> mutations in the  $\beta$ -subunit are adjacent to Fork Loop 2 (FL2) which has been observed to adopt multiple conformations, one where it interacts with the leading

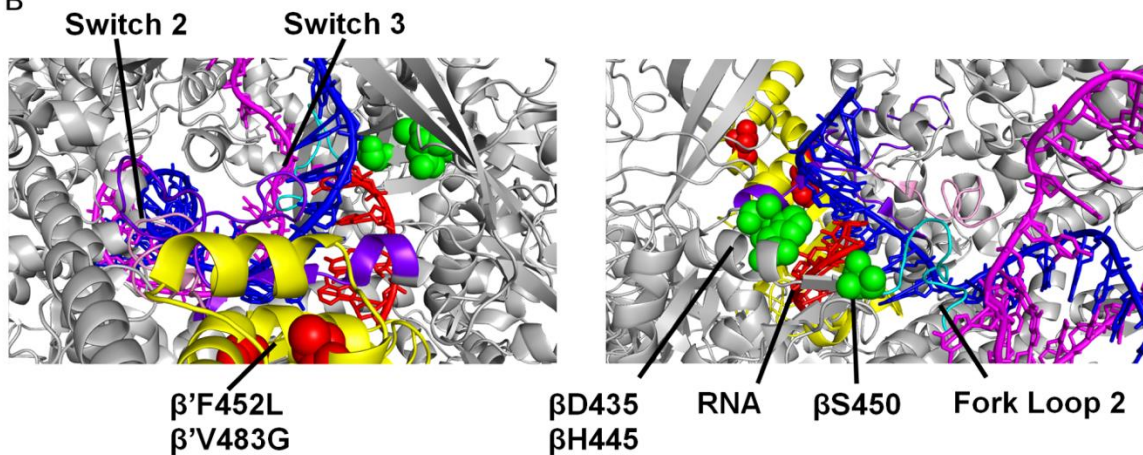
edge of the transcription bubble and others where it makes contacts with the bridge helix<sup>23-24</sup>. In the *E. coli* RNAP crystal structures with RMP, FL2 of the  $\beta$ S531L ( $\beta$ S450L in MTB numbering) mutant was disordered, reducing contacts with RMP (likely the source of the mutant RNAP's ~10,000-fold reduction in affinity for RMP). A disordered FL2 was not observed in the WT *E. coli* RMP complex structure<sup>9</sup>. This structural difference suggests that a defect in the conformational stability of FL2 is likely a contributor to the destabilization of the RPo observed in the RMP<sup>R</sup> mutants.

Interestingly both compensatory mutations studied,  $\beta$ 'F452L and  $\beta$ 'V483G, had a stabilizing effect on RPo half-life for the  $\beta$ S450L mutant. The degree of stabilization for the  $\beta$ 'F452 and  $\beta$ 'V483G mutations partially and fully complemented the defect observed for the  $\beta$ S450L mutation; consistent with growth rate and fitness determination measurements of *M. smegmatis* expressing recombinant MTB *rpoB-rpoC* alleles with these mutations<sup>18</sup>. The switch region of RNAP aids in the closure of the RNAP clamp on DNA and forms a complex network which directly interacts with the template strand (see **Figure III-5**)<sup>25-26</sup>. Mutations in the switch region have been shown to negatively affect open-promoter complex stability via weakening the strength of the RNAP clamp on the DNA<sup>26-29</sup>. Switch 3 makes extensive contacts with the alpha-helical region above the DPBB where the compensatory mutations studied here reside and Switch 2 feeds into  $\beta$ -sheet 1 of the  $\beta$ -barrel. Compensatory mutations in the DPBB domain may induce conformational changes in the switch region to rectify defects in RPo stability caused by the RMP<sup>R</sup> mutations in the  $\beta$ -subunit through interaction with the template strand (see **Figure III-5A** and **Figure III-5B** left panel).

A



B



**Figure III-5: Structure of *Mycobacterium smegmatis* RNAP with full transcription bubble (PDB: 5VI5) <sup>30</sup>.** **A)** Highlighted structural motifs which contribute to effects observed on transcription. **B)** Perspective of RNA•DNA heteroduplex region from the perspective of the compensatory mutations in the DPBB (left panel) and the Rif<sup>R</sup> mutations (right panel).

Surprisingly, despite being the most prevalent of the RMP<sup>R</sup> mutants present in isolates, the  $\beta$ S450L mutant showed the greatest defect in its elongation rate, whereas the  $\beta$ D435V and  $\beta$ H445Y RMP<sup>R</sup> mutants increased the rate of elongation relative to WT (see **Figure III-2**). The variability in rate of elongation among the RMP<sup>R</sup> mutants is curious, but not unprecedented. Mutations in this region of the  $\beta$ -subunit have been shown to have varying effects on elongation rate <sup>10</sup>. A laboratory-evolved RMP<sup>R</sup> mutation in the *E. coli* RNAP,  $\beta$ I572N, resulted in an attenuated elongation rate while mutation of

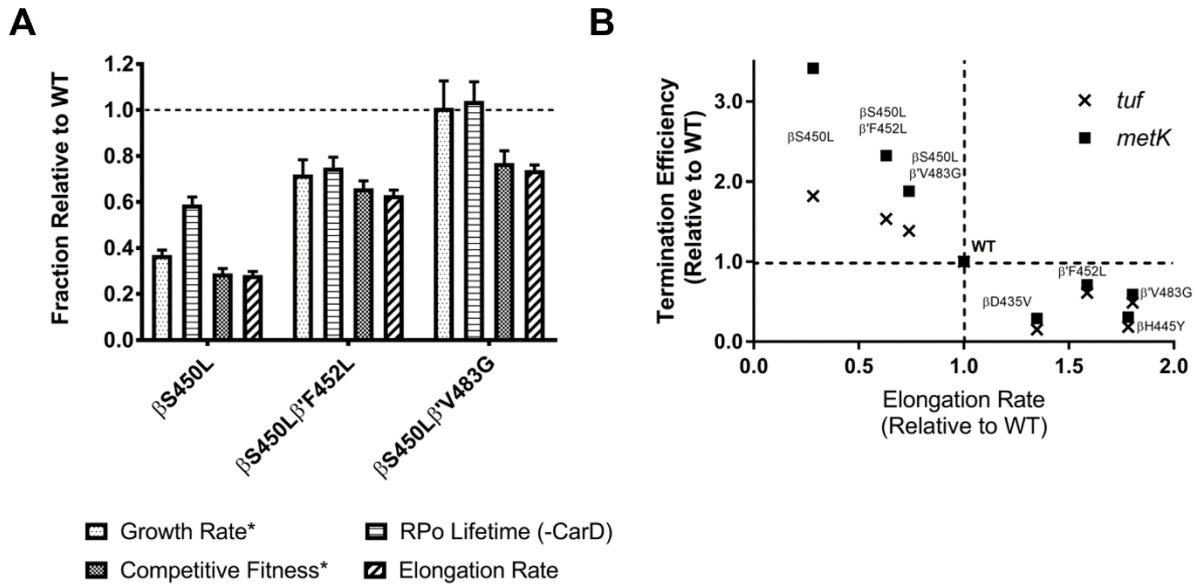
the same amino acid to a threonine had no effect on elongation rate <sup>10</sup>. There are other instances in which mutations in this region of the  $\beta$ -subunit effect elongation rate (increases and decreases relative to WT) <sup>10, 31-35</sup>. The  $\beta$ S450L mutation is the closest of the three RMP<sup>R</sup> mutations to FL2, which does adopt multiple conformations, suggesting it plays a role in the elongation <sup>23-24, 36</sup>. Additionally, FL2 stimulates elongation rate via sequestration of incoming NTP and promotion of phosphodiester bond formation. It has been suggested this occurs by allosterically influencing fork loop1 (FL1) <sup>37</sup>. FL1 makes up part of the active site and has a glutamate which may coordinate NTPs in the  $i+1$  site <sup>37</sup>. The  $\beta$ S450L mutation may indirectly affect the function of FL1 in the sequestration of NTP substrate, attenuating the elongation rate.

Conformational changes in the fork domain (fork loop 1 and fork loop 2) are thought to promote RNAP clamp closure around the RNA•DNA heteroduplex, helping to maintain elongation complex stability <sup>26</sup>. The RMP<sup>R</sup> mutations in the fork domain have been shown to cause slippage along homopolymeric template DNA with *E. coli* and *S. cerevisiae* RNAP <sup>10, 38-39</sup>. It is thought that this occurs from weakening of the clamp causing spontaneous melting of the heteroduplex region <sup>10</sup>. The RMP<sup>R</sup> mutations in this study are clustered around this heteroduplex region on the RNA side of the RNA•DNA heteroduplex (see **Figure III-5**). Mutations in this pocket affect the steric and surface (polarity/electrostatic) parameters of the RNA exit channel and have the potential for either weakening or strengthening interactions between the residues within the rifamycin-binding cleft and the extending nascent RNA <sup>9</sup>. As one consequence, mutations in the RMP cleft could have significant effects on translocation of the RNAP and elongation rate. Given the proximity of the RMP<sup>R</sup> mutations to the DFDGD motif, an alternative

explanation for these effects is (perhaps subtle) alteration of the catalytic magnesium position (see **Figure I-5**).

“Backtracking” is a process where the RNA•DNA hybrid region translocates in the reverse direction, shifting the 3'- end of the RNA into the *i* site and re-forming the post-incorporation/pre-translocation state. This most often happens in the process of “proof-reading”<sup>40-41</sup>. Weakened RNA•DNA hybrid regions are associated with an increased ability for the heteroduplex to backtrack<sup>42-43</sup>, preventing incorporation of any new NTP and rendering the 3' nucleotide susceptible to hydrolysis. We observed that the mutations studied have varying effects on the hydrolysis of the 3' nucleotide of the RNA8 (see **Figure III-4**). Consistent with observations for RNAPs which are deficient in backtracking (faster RNAPs are less prone to backtracking), diminished exonuclease activity was observed for the  $\beta$ D435V and  $\beta$ H445Y mutants<sup>42</sup>. In contrast, the  $\beta$ S450L/ $\beta$ 'V483G mutation dramatically increases hydrolysis of 3'-UMP from RNA8 while also decreasing the elongation rate. The RMP<sup>R</sup> mutations (on the RNA side of the RNA•DNA hybrid) may affect the RNA•DNA hybrid such that forward translocation is increased, resulting in increased elongation rates and reduced backtracking and hydrolysis. While the compensatory mutations (which occur on the DNA side of the heteroduplex) may have similar effects on the RNA•DNA hybrid, they could also perturb the region around the catalytic magnesium, where changes in spatial environment have been shown enhance 3'-5' exonuclease activity<sup>43</sup>.

RNAPs which do not efficiently recognize termination signals will continue transcribing neighboring genes. Screening performed to identify altered termination phenotypes in *E. coli* found mutations enriched near fork loop 2 and the RMP binding site



**Figure III-6: Concluding observations from kinetic studies of MTB RNAPs. A)** Similarity between Open-promoter complex and Growth as well as elongation rate and competitive fitness observed for the clinically relevant genotypes (\*Competitive fitness and growth were determined in Song et al. <sup>18</sup>). **B)** Inverse correlation between elongation rate and termination efficiency observed for the MTB RNAP variants in this study.

<sup>12</sup>. Our observations indicate that mutant RNAPs carrying the  $\beta$ S450L are not just more efficient at intrinsic termination than the other RMP<sup>R</sup> mutants studied but are also more efficient than the WT RNAP. Although the presence of a single compensatory mutation yields a slight decrease in termination efficiency, the double mutants maintain termination efficiencies which are still greater than WT at both the *tuf* and *metK* terminators (see **Figure III-3**).

Interestingly, as noted above for RPo stability, the relative elongation rates for the mutants containing the S450L in this study also mirror the relative competitive fitness observed in *M. smegmatis* (see **Figure III-6A**).<sup>18</sup> Production of rRNA in vivo is coupled to

growth rate and organism fitness.<sup>44</sup> It is possible that slow-growing bacteria such as *M. tuberculosis*, which has an elongation rate 15 times slower than that of *E. coli*, could present fitness defects that are reflective of reduced production of rRNA. We also observed that elongation rate and termination efficiency are inversely correlated (see **Figure III-6B**), which is consistent with previous reports<sup>16, 33</sup>. Our results suggest that this aspect of transcription may be the driving factor for initial RMP<sup>R</sup> mutant selection. While many factors contribute to the fitness of the organism, these intrinsic properties of the RMP<sup>R</sup> mutants will likely have a major impact on fitness.

Trans-acting Nus factors A and G from MTB were also tested to determine the extent to which they alter termination efficiencies (see **Figure III-4**). As previously reported, MTB NusA had no effect on the termination efficiencies, despite observed effects in an *E. coli*-based transcription system. However, MTB NusG did increase the termination efficiencies for many of the mutant RNAPs tested, though not all. It had limited effects on termination when termination efficiencies were very low or very high. For low termination efficiency RNAPs, the elongation rate may be too fast for NusG to enhance termination and the incremental effect of NusG may be masked in high termination efficiency RNAPs.

Our study explored the molecular basis for the fitness defects of certain RMP<sup>R</sup> mutants observed in MTB clinical isolates. We found that there are distinct mechanistic and kinetic differences for each RMP<sup>R</sup> mutant. These studies provide insight into the mechanistic framework for the evolution of secondary, compensatory mutations observed in the  $\beta'$ -subunit of RNA polymerase. We conclude that a critical factor is the balance between elongation rate and termination efficiency and that this likely explains the



prevalence of the  $\beta$ S450L mutation and contributes to the presence of compensatory mutations. Another factor which seems to drive secondary mutations in the  $\beta'$ -subunit is their effect on RPo stability. Further studies which probe the structural consequences of the RMP<sup>R</sup> mutant fitness defect compensation by compensatory mutations in the  $\beta'$ -subunit are needed to confirm the exact mechanism of action of these mutations.

### **Notes to Chapter III**

This work was supported by NIH grant R01-AI110780 (GAG) and the University of Michigan, College of Pharmacy. We would also like to acknowledge Prof. Katsu Murakami and Catherine Sutherland (Penn. State Univ.) and members of the Garcia lab for critical reading and helpful suggestions about the manuscript. I would like to thank Fatima Ugur for producing the *rpoC* gene with the F452L mutation. Parts of this chapter are published in Stefan et al, *Antimicrobial Agents and Chemotherapy*, **2018**, DOI: 10.1128/AAC.00164-18.

## References

1. Andersson, D. I.; Levin, B. R., The biological cost of antibiotic resistance. *Curr Opin Microbiol* **1999**, 2 (5), 489-93.
2. Sundqvist, M., Reversibility of antibiotic resistance. *Ups J Med Sci* **2014**, 119 (2), 142-8.
3. Gagneux, S.; Long, C. D.; Small, P. M.; Van, T.; Schoolnik, G. K.; Bohannon, B. J., The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **2006**, 312 (5782), 1944-6.
4. Billington, O. J.; McHugh, T. D.; Gillespie, S. H., Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **1999**, 43 (8), 1866-9.
5. Mariam, D. H.; Mengistu, Y.; Hoffner, S. E.; Andersson, D. I., Effect of *rpoB* mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **2004**, 48 (4), 1289-94.
6. Gagneux, S., Fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Clin Microbiol Infect* **2009**, 15 Suppl 1, 66-8.
7. Gygli, S. M.; Borrell, S.; Trauner, A.; Gagneux, S., Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. *FEMS Microbiol Rev* **2017**, 41 (3), 354-373.
8. Comas, I.; Borrell, S.; Roetzer, A.; Rose, G.; Malla, B.; Kato-Maeda, M.; Galagan, J.; Niemann, S.; Gagneux, S., Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature genetics* **2012**, 44 (1), 106-10.
9. Molodtsov, V.; Scharf, N. T.; Stefan, M. A.; Garcia, G. A.; Murakami, K. S., Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most

clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Mol Microbiol* **2017**, *103* (6), 1034-1045.

10. Zhou, Y. N.; Lubkowska, L.; Hui, M.; Court, C.; Chen, S.; Court, D. L.; Strathern, J.; Jin, D. J.; Kashlev, M., Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem* **2013**, *288* (4), 2700-10.

11. Zhou, Y. N.; Jin, D. J., The rpoB mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proc Natl Acad Sci U S A* **1998**, *95* (6), 2908-13.

12. Jin, D. J.; Walter, W. A.; Gross, C. A., Characterization of the termination phenotypes of rifampicin-resistant mutants. *J Mol Biol* **1988**, *202* (2), 245-53.

13. McDowell, J. C.; Roberts, J. W.; Jin, D. J.; Gross, C., Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. *Science* **1994**, *266* (5186), 822-5.

14. Mitra, A.; Angamuthu, K.; Nagaraja, V., Genome-wide analysis of the intrinsic terminators of transcription across the genus *Mycobacterium*. *Tuberculosis (Edinb)* **2008**, *88* (6), 566-75.

15. Temiakov, D.; Anikin, M.; McAllister, W. T., Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J Biol Chem* **2002**, *277* (49), 47035-43.

16. Czyz, A.; Mooney, R. A.; Iaconi, A.; Landick, R., Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* **2014**, *5* (2), e00931.

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

18. Song, T.; Park, Y.; Shamputa, I. C.; Seo, S.; Lee, S. Y.; Jeon, H. S.; Choi, H.; Lee, M.; Glynn, R. J.; Barnes, S. W.; Walker, J. R.; Batalov, S.; Yusim, K.; Feng, S.; Tung, C. S.; Theiler, J.; Via, L. E.; Boshoff, H. I.; Murakami, K. S.; Korber, B.; Barry, C. E., 3rd; Cho, S. N., Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis*

are amplified under conditions of nutrient starvation and compensated by mutation in the beta' subunit of RNA polymerase. *Mol Microbiol* **2014**, *91* (6), 1106-19.

19. Mondal, S.; Yakhnin, A. V.; Sebastian, A.; Albert, I.; Babitzke, P., NusA-dependent transcription termination prevents misregulation of global gene expression. *Nature Microbiology* **2016**, *1*, 1-8.
20. Kalyani, B. S.; Kunamneni, R.; Wal, M.; Ranjan, A.; Sen, R., A NusG paralogue from Mycobacterium tuberculosis, Rv0639, has evolved to interact with ribosomal protein S10 (Rv0700) but not to function as a transcription elongation-termination factor. *Microbiology* **2015**, *161* (Pt 1), 67-83.
21. Schmidt, M. C.; Chamberlin, M. J., nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J Mol Biol* **1987**, *195* (4), 809-18.
22. Yakhnin, A. V.; Babitzke, P., NusA-stimulated RNA polymerase pausing and termination participates in the Bacillus subtilis trp operon attenuation mechanism invitro. *Proc Natl Acad Sci U S A* **2002**, *99* (17), 11067-72.
23. Kettenberger, H.; Armache, K. J.; Cramer, P., Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **2004**, *16* (6), 955-65.
24. Meyer, P. A.; Ye, P.; Suh, M. H.; Zhang, M.; Fu, J., Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data. *J Biol Chem* **2009**, *284* (19), 12933-9.
25. Gnatt, A. L.; Cramer, P.; Fu, J.; Bushnell, D. A.; Kornberg, R. D., Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **2001**, *292*.
26. Pupov, D.; Miropolskaya, N.; Sevostyanova, A.; Bass, I.; Artsimovitch, I.; Kulbachinskiy, A., Multiple roles of the RNA polymerase {beta}' SW2 region in transcription initiation, promoter escape, and RNA elongation. *Nucleic Acids Res* **2010**, *38* (17), 5784-96.
27. Rutherford, S. T.; Villers, C. L.; Lee, J. H.; Ross, W.; Gourse, R. L., Allosteric control of Escherichia coli rRNA promoter complexes by DksA. *Genes Dev* **2009**, *23* (2), 236-48.

28. Wiesler, S. C.; Burrows, P. C.; Buck, M., A dual switch controls bacterial enhancer-dependent transcription. *Nucleic Acids Res* **2012**, *40* (21), 10878-92.
29. Zhang, N.; Schafer, J.; Sharma, A.; Rayner, L.; Zhang, X.; Tuma, R.; Stockley, P.; Buck, M., Mutations in RNA Polymerase Bridge Helix and Switch Regions Affect Active-Site Networks and Transcript-Assisted Hydrolysis. *J Mol Biol* **2015**, *427* (22), 3516-26.
30. Hubin, E. A.; Lilic, M.; Darst, S. A.; Campbell, E. A., Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nat Commun* **2017**, *8*, 16072.
31. Petushkov, I.; Pupov, D.; Bass, I.; Kulbachinskiy, A., Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Res* **2015**, *43* (12), 5798-809.
32. Tavormina, P. L.; Reznikoff, W. S.; Gross, C. A., Identifying interacting regions in the beta subunit of Escherichia coli RNA polymerase. *J Mol Biol* **1996**, *258* (2), 213-23.
33. Tavormina, P. L.; Landick, R.; Gross, C. A., Isolation, purification, and in vitro characterization of recessive-lethal-mutant RNA polymerases from Escherichia coli. *J Bacteriol* **1996**, *178* (17), 5263-71.
34. Jin, D. J.; Gross, C. A., RpoB8, a rifampicin-resistant termination-proficient RNA polymerase, has an increased Km for purine nucleotides during transcription elongation. *J Biol Chem* **1991**, *266* (22), 14478-85.
35. Heisler, L. M.; Suzuki, H.; Landick, R.; Gross, C. A., Four contiguous amino acids define the target for streptolydigin resistance in the beta subunit of Escherichia coli RNA polymerase. *J Biol Chem* **1993**, *268* (34), 25369-75.
36. Vassylyev, D. G.; Vassylyeva, M. N.; Perederina, A.; Tahirov, T. H.; Artsimovitch, I., Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **2007**, *448* (7150), 157-62.
37. Kireeva, M. L.; Domecq, C.; Coulombe, B.; Burton, Z. F.; Kashlev, M., Interaction of RNA polymerase II fork loop 2 with downstream non-template DNA regulates transcription elongation. *J Biol Chem* **2011**, *286* (35), 30898-910.

38. Strathern, J.; Malagon, F.; Irvin, J.; Gotte, D.; Shafer, B.; Kireeva, M.; Lubkowska, L.; Jin, D. J.; Kashlev, M., The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in *S. cerevisiae*. *J Biol Chem* **2013**, *288* (4), 2689-99.
39. Liu, X.; Martin, C. T., Transcription elongation complex stability: the topological lock. *J Biol Chem* **2009**, *284* (52), 36262-70.
40. Bar-Nahum, G.; Epshtein, V.; Ruckenstein, A. E.; Rafikov, R.; Mustaev, A.; Nudler, E., A ratchet mechanism of transcription elongation and its control. *Cell* **2005**, *120* (2), 183-93.
41. Kireeva, M. L.; Nedialkov, Y. A.; Cremona, G. H.; Purtov, Y. A.; Lubkowska, L.; Malagon, F.; Burton, Z. F.; Strathern, J. N.; Kashlev, M., Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell* **2008**, *30* (5), 557-66.
42. Nudler, E.; Mustaev, A.; Lukhtanov, E.; Goldfarb, A., The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **1997**, *89* (1), 33-41.
43. Sosunov, V.; Sosunova, E.; Mustaev, A.; Bass, I.; Nikiforov, V.; Goldfarb, A., Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J* **2003**, *22* (9), 2234-44.
44. Zhu, M.; Dai, X., On the intrinsic constraint of bacterial growth rate: M. tuberculosis's view of the protein translation capacity. *Crit Rev Microbiol* **2018**, 1-10.

### Chapter III Appendix

Primer	Sequence	Purpose
<b>Mt rrnA3 -66 AMP FOR</b>	GATCACATGTATGGATATCTATGGATGACCGAACCTGG	Amplification of rrnAP3 -66 from TSS
<b>Mt rrnA3 +351 AMP REV</b>	GATCGAATTCCCAGTTTCCCAGGCTTATCCCGAAGTGC	Amplification of rrnAP3 +351 from TSS
<b>SynB Terminator FOR</b>	GACTTCTAGAAGAAAAAAAAAAGCGCCGCAACTGCGGCGCTTTT TTTTTTCAGGTATCTATCTAGAGACT	Synthetic terminator for <i>in vitro</i> transcription
<b>SynB Terminator REV</b>	AGTCTCTAGATAGATACCTGAAAAAAAAAAGCGCCGCAGTTGCG GCGCTTTTTTTTTTCTTCTAGAAGTC	Synthetic terminator for <i>in vitro</i> transcription
<b>EC26 DNA Fragment FOR</b>	GACATGTAAATATTTGTTGTTAACTCTTGACAAAAGTGTTAAAAGC GGCTAGTATTTAAAGGGATGGATGAGATTTGAAGGTTGGGTCCC ATGGCCAGAT	Consensus <i>E. coli</i> promoter for termination studies
<b>EC26 DNA Fragment REV</b>	ATCTGGCCATGGGACCCAACCTTCAAATCTCATCCATCCCTTTAA ATACTAGCCGCTTTTAACTTTTGTCAAGAGTTAACAACAAATAT TTACATGTC	Consensus <i>E. coli</i> promoter for termination studies
<b>EC19 AMP FOR</b>	ATGGATATCTATGGATGACCGAACCTGG	Amplification of DNA template
<b>EC26 AMP FOR</b>	TTGTTGTTAACTCTTGACAAAAGTGTTAAAAGCGG	Amplification of DNA template
<b>EC AMP REV</b>	GCCTGCAGGTTCGACTCTAGAGG	Amplification of DNA template
<b>rrnA3 EC19 FOR</b>	GACTGGCAGGGTTGCCCGTTGCGGGC0GG	Stalled Elongation Mutagenesis
<b>rrnA3 EC19 REV</b>	CCGCCCGCAACGGGGCAACCCTGCCAGTC	Stalled Elongation Mutagenesis
<b>CarD AMP FOR</b>	GGGAATTCCATATGATTTTCAAGGTCGGAGACACCGTTGTC	Amplification of CarD from gDNA†
<b>CarD AMP REV</b>	GATCCC <b>GGATCCT</b> CAAGACGCGGCGGCTAAAACCTCGTCAAG	Amplification of CarD from gDNA†
<b>NusA AMP FOR</b>	CCGG <b>CATATG</b> AACATCGACATGGCTGCTCTGCATGCC	Amplification of NusA from gDNA†
<b>NusA AMP REV</b>	ACGCG <b>GATCCT</b> TAGCGGTCTGCGCCATACCGC	Amplification of NusA from gDNA†
<b>NusG AMP FOR</b>	GCGCCATATGGTGACTACCTTCGACGGT	Amplification of NusG from gDNA†
<b>NusG AMP REV</b>	GCGCCTCGAGCTAGATCTTGGGAGACTTG	Amplification of NusG from gDNA†

- Restriction sites are in bold and mutagenesis sequences are underlined. †: H37Rv genomic DNA (gDNA)

Appendix Table III-A1: Primers used in Chapter III

<b>Plasmid</b>	<b>Purpose</b>
pET19bpps-CarD	Expression of MTB CarD
pET19bpps-NusA	Expression of MTB NusA
pET19bpps-NusG	Expression of MTB NusG
pMt-rrnA3	Template manipulation vector for assays
pMt-rrnA3-synB	Open-promoter half-life determination
pEC19	Amplification of EC19 template
pEC26-Tuf	Amplification of Tuf terminator containing template
pEC26-MetK	Amplification of MetK terminator containing template

**Appendix Table III-A2: Plasmids used in Chapter III**

	<b>% Active RNAP</b>
WT	55
$\beta$ D435V	73
$\beta$ H445Y	49
$\beta$ S450L	53
$\beta$ 'F452L	67
$\beta$ 'V483G	64
$\beta$ S450L $\beta$ 'F452L	47
$\beta$ S450L $\beta$ 'V483G	47

**Appendix Table III-A3: Percent of active RNAP from each preparation as determined from primer extension.** Percent activity was calculated by dividing the sum of density for the 7-mer and 9-mer by the total density in the lane for each RNAP. Values are averages derived from three independent measurements.

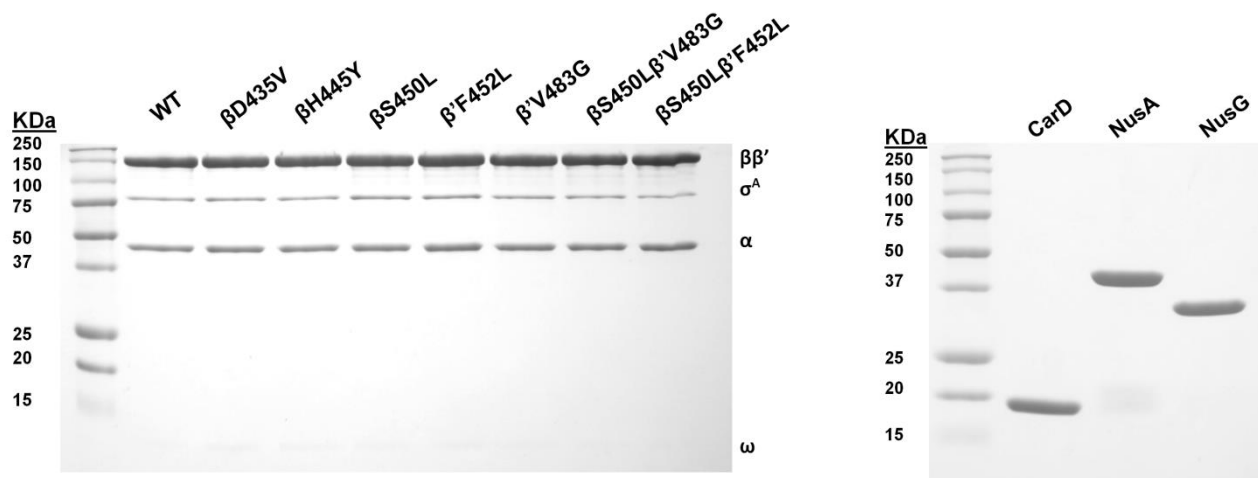


-66  
 ATGGATATCTATGGATGACCGAACCTGGTCTT**GACT**CCATTGCCGGATTTGTATT**AGACT**GGCAGGGTTGCCCGAAGCGGGCGGAAACAAGCAA  
 GCGTGTGTTTGAGAACTCAATAGTGTGTTGGTGGTTTCACATTTTGTGTTATTTTTGGCCATGCTCTTGATGCCCCGTTGTCGGGGCGTGGC  
 CGTTTGTGTTGTCAGGATATTTCTAAATACCTTTGGCTCCCTTTTCCAAAGGAGTGTTGGGTTTTGTTGGAGAGTTTGATCCTGGCTCAGGACGA  
 ACCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGTCTCTTCGGAGATACTCGAGTGCGAACGGGTGAGTAACACGTGGGTGATCTG  
 CCCTGCACTTCGGGATAAGCCTGGGAACTGGGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGC  
 +398

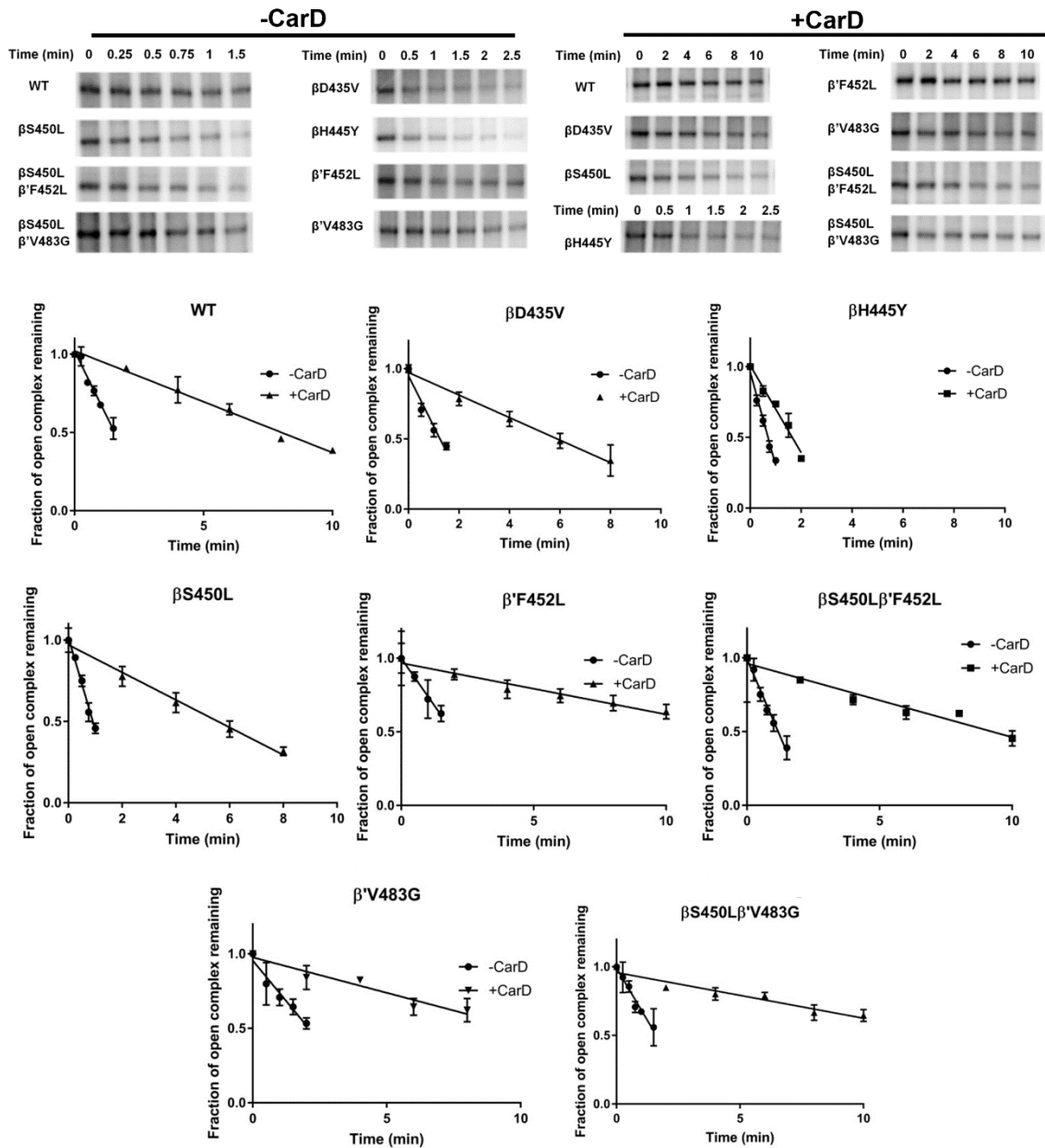
pMt-EC19  
 ↑↑  
 TT  
 EC19  
 ↓

**Appendix Figure III-A1: Sequence of source DNA template used in Chapter III.**

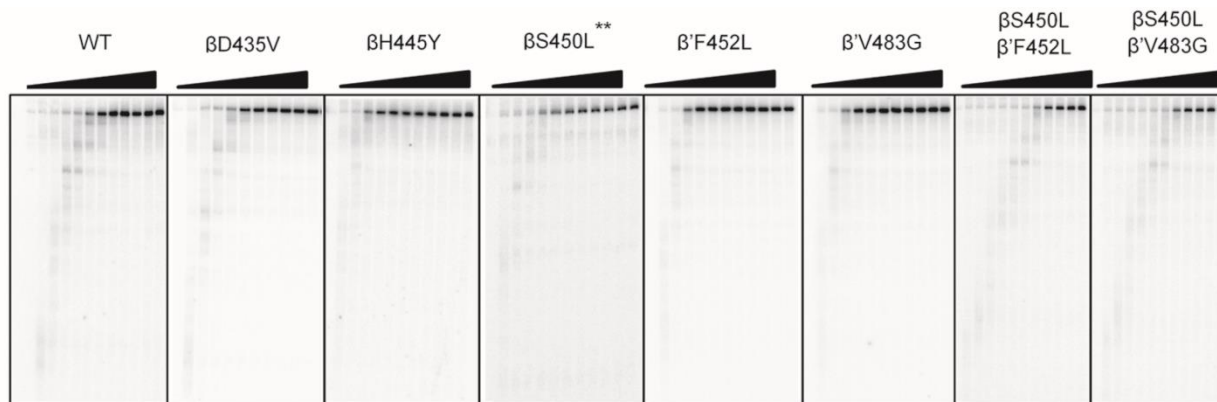
Sequence is of native H37Rv genomic DNA (+1468189 to +1468605) -66 to +351 of the transcription start site of the *rmA3* promoter. The -10 and -35 elements are highlighted and the transcription start site is in green. Sites of mutagenesis are also highlighted. Endogenous *MscI* restriction site used for pEC26 plasmid construction is not highlighted.



**Appendix Figure III-A2: SDS-PAGE of all proteins used in Chapter III. RNAPs (left panel) and trans-acting transcription factors (right panel).**

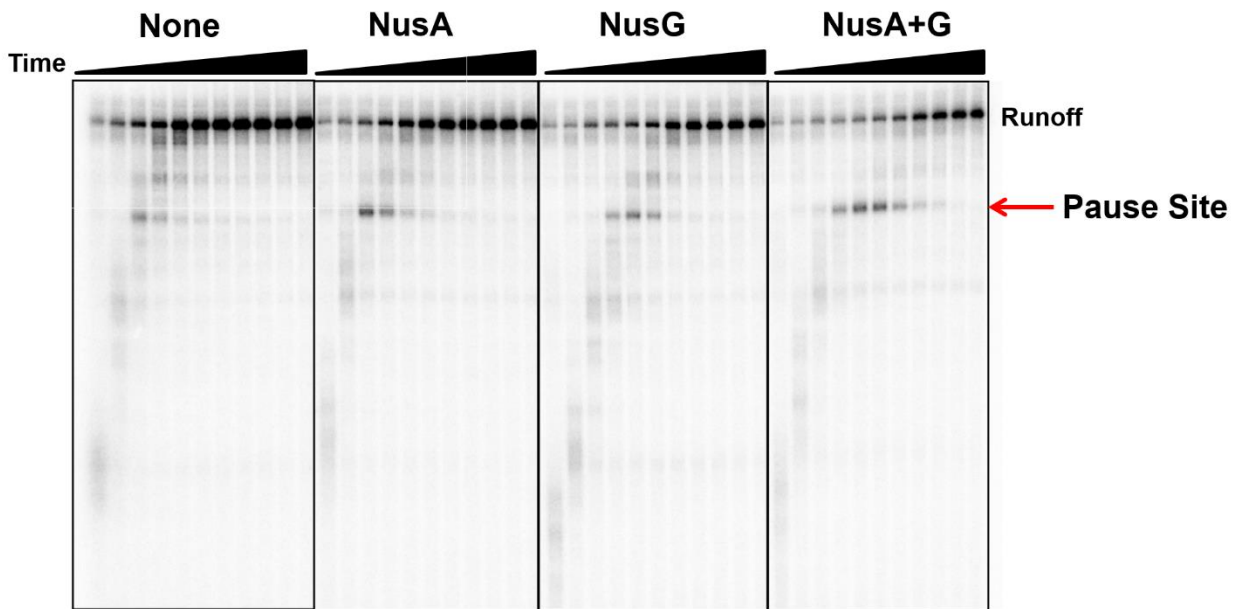


**Appendix Figure III-A3: Gel images and plots for determined open-promoter complex half-life.** **A)** Gel images used to determine open-complex stability for each RNAP in the presence and absence of CarD. **B)** Plots showing the decay of the open-promoter complex over time for each of the RNAPs in this study. All experiments are performed in triplicate.



**Appendix Figure III-A4: Plots used to determine the elongation rate for MTB RNAPs.**

Density of runoff transcript was plotted against time. The time at which 50% of elongating RNAPs reach runoff length was divided by the length of the transcribed region to determine the average elongation rate. Phosphorimage of elongating MTB RNAPs. Time points for all RNAPs except the  $\beta$ S450L mutant (min): 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 5, 15. Time points for the  $\beta$ S450L mutant (min): 2, 3, 4, 5, 7.5, 10, 12.5, 15, 20, 30, 45.



**Appendix Figure III-A5: Effect of NusA and NusG on elongation rate with WT MTB RNAP.** Time points are as follows: 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 5, 15 minutes.

## CHAPTER IV

### High-throughput Screening to Identify Inhibitors of the *M. tuberculosis* Transcriptional Regulator CarD•RNAP Interaction

As mentioned in previous chapters RNA polymerase represents an attractive therapeutic target, when its function is compromised it is proven to be effective in bacterial killing. Several bacterial RNAP inhibitors have been identified over the last few decades (see **Figure I-7**); however, only two have made it to the clinic, Rifampin and Fidaxomicin. Unfortunately, resistance to many of these inhibitors has been reported when MTB is challenged with the compounds. Traditionally, most of the work in identifying methods of targeting RNAP have focused on the enzyme activity; however, RNAP functionality *in vivo* is significantly more complex, requiring several trans-acting factors which are essential for proper gene regulation and viability.<sup>1-2</sup>

In the case of MTB, several transcription factors have been shown to be required for proper transcription and organism viability, one of these factors is the global transcriptional regulator CarD.<sup>1, 3</sup> Mycobacteria form relatively unstable open-promoter complexes (RPO) where promoter DNA is unwound. CarD is required for stabilization of these complexes.<sup>4-5</sup> CarD exists in two subdomains, its N-terminal domain (1-53) interacts with MTB RNAP at the  $\beta$ 1 $\beta$ 2-lobes of the  $\beta$ -subunit, also known as the “protrusion”. Its C-terminal domain (64-162), which is separated from the N-terminal domain by a 10-amino acid linker, interacts with promoter DNA at the upstream fork junction.<sup>6-7</sup> CarD is required for MTB viability and mediation of stress responses such as exposure to antibiotics and

oxidative stress.<sup>1, 3</sup> CarD's function as a global transcriptional regulator that is required for MTB survival makes it a potential novel therapeutic target.

In this chapter, the development, optimization, and the validation of a fluorescence polarization assay to monitor the interaction between CarD and the RPo is presented. A pilot screen (237 compounds) was conducted to determine the viability of this assay for high throughput screening (HTS) prior to performing a 23,320 small molecule HTS. Hits from this screen were characterized in biochemical, biophysical, and bacterial assays for validation and to probe their mechanism(s) of action.

## Materials and Methods

### *Protein Manipulation, Purification, and Labeling*

*Wild-type* CarD was prepared as described in Chapter III and MTB RNAP was prepared as described in Chapter II. Mutant CarD variants were produced by 2-step mutagenesis using primers described in **Appendix Table IV-A1**. For fluorophore-labeled CarD, BL21(DE3) *E. coli* transformed with each pET19bbps-CarD plasmid was grown to an OD<sub>600</sub> of 0.6 in 2xTY and induced with 1mM IPTG for 4 hours at 28°C. Cells were harvested and pellets from 1L of culture were resuspended in 20 mL of Lysis Buffer (20 mM Tris HCl (pH 8.0), 500mM NaCl, 0.1mM EDTA, 5% glycerol (v/v), 5mM β-mercaptoethanol (BME)). Before lysis, the suspension was supplemented with 2 mM PMSF, 1X Roche Protease cocktail, 1 mg/mL lysozyme, and 200 U DNaseI. Cells were disrupted by sonication and subsequently pelleted via centrifugation at 30,000 x g for 45 minutes at 4 °C. Clarified lysate was sterile filtered and applied to a 1 mL HisTrap HP column. CarD was eluted by step gradient to 500 mM imidazole after extensive washing.

To remove imidazole and  $\beta$ ME, CarD was precipitated with 0.3 g/mL of ammonium sulfate followed by rinsing the protein pellet with Labeling Buffer (100 mM Tris HCl (pH 7.5 at 25°C), 100 mM NaCl, and 1mM EDTA) with the same concentration of ammonium sulfate after which the protein was pelleted via centrifugation for 30 min at 3,000 x g. Precipitated CarD was resuspended in Labeling buffer to 2 mg/mL or until the solution was clear. To ensure all cysteines were reduced a 10-fold molar excess of freshly made TCEP solution was added and allowed to incubate at 25°C for 1 hour prior to labeling. In complete darkness 10-fold molar excess BODIPY™ FL iodoacetamide (N-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-yl)methyl)iodoacetamide, Thermo Fisher Scientific) suspended in dry DMF was added dropwise to the CarD solution with stirring. The reaction was allowed to proceed for 2.5 hours before excess  $\beta$ ME was added to quench the reaction. The reaction was diluted in lysis buffer and reapplied to 1 mL of NTA-Ni<sup>2+</sup> agarose, where it was allowed to incubate for 1 hour. The applied solution was allowed to flow through and the resin was washed with 5 mL of lysis buffer, then eluted with 5 mL of lysis buffer with 500 mM imidazole. Precision protease was added to the solution which was then dialyzed overnight in SEC Buffer (20 mM Tris HCl (pH8.0), 300mM NaCl, 0.1mM EDTA, 5% glycerol, and 2 mM DTT).

To remove the Precision protease, the dialyzed protein was allowed to flow through 1 mL of glutathione agarose before being concentrated to ~0.5mL followed by application to a Superdex 200 10/300 GL column equilibrated with SEC buffer. Fractions containing labeled CarD were concentrated to >1 mg/mL, flash frozen with N<sub>2(l)</sub> and stored at -80°C. The extents of labeling of the dye-conjugated CarD proteins were confirmed both spectroscopically using **Equation (1)**, where  $A_{504}$  is the absorbance of sample at

504 nm and  $\epsilon_{504}$  is the extinction coefficient for BODIPY™ FL (68,000 M<sup>-1</sup>•cm<sup>-1</sup>) as well as by LCMS (see **Appendix Table IV-A2** and **Figure IV-1-9**). Labeling efficiencies ranged from 40-100% and are described in **Appendix Table IV-A2**.

$$\text{Equation (1)} \quad \frac{A_{504}}{\epsilon_{504}} \times \frac{\text{MW of CarD}}{\text{CarD mg/mL}} = \frac{\text{moles of dye}}{\text{moles of CarD}}$$

RNA polymerase binding protein (RbpA) from *Mycobacterium smegmatis* was purified in a similar fashion to that described by Dey et al.<sup>8</sup> In brief, pETMsRbpA transformed BL21(DE3) cells were grown to an OD<sub>600</sub> of 0.6 after which cells were induced at 30°C for 4 hours with 1 mM IPTG. Cells from 1 L of culture were resuspended in 20 mL of lysis buffer and disrupted by sonication. Clarified lysate was filtered and applied to a 1 mL HisTrap HP column and eluted with a 10-column volume linear gradient of 20 to 500 mM imidazole in lysis buffer. Fractions with RbpA were concentrated and applied to a Superdex 200 10/300 GL column equilibrated with SEC buffer. Purified RbpA was concentrated to 4 mg/mL, frozen in N<sub>2(l)</sub>, and stored at -80°C.

#### *CarD Fluorescence Polarization Assay*

The fluorescence polarization (FP) assay was conducted as described below in a final volume of 20  $\mu$ L. Optimized CarD FP assay buffer contains 20 mM Tris HCl (pH 7.9 at 37°C), 150mM potassium glutamate (KGlu), 10 mM MgCl<sub>2</sub>, 0.1mM EDTA, 0.01% Triton X-100, 5mM DTT unless noted otherwise. For binding isotherms used to determine dissociation constants (K<sub>D</sub>), BODIPY-labeled CarD tested at 1 nM, 3 nM and 5 nM, rrnAP3 and artificial bubble DNA fragments (the latter of which has a 12 base noncomplementary region from -11 to +1 of the TSS, **Appendix Table IV-A1**) were kept constant at 1  $\mu$ M



when used. WT MTB RNAP was typically titrated from 1  $\mu$ M down to 244 pM. Assay components were allowed to incubate in a humidified incubator at 37°C for 2 hours in a black Corning® low volume round-bottom non-binding surface 384-well plate before reading in a BioTek Synergy H1 Hybrid Multi-mode plate reader preheated to 37°C. Wells were read at a sensitivity of 100 with 15 flashes per well at an excitation of 485/20 nm and emission of 528/20 nm. Fluorescence polarization (P) was determined by subtracting a blank well containing all assay components except BODIPY-labeled CarD from both the parallel ( $F_{\parallel}$ ) and perpendicular ( $F_{\perp}$ ) intensities and **Equation (2)** was used to determine polarization. FP values for binding isotherms were fit to a sigmoidal dose-response with variable slope described in **Equation (3)**, where Y = millipolarization (mP) and X = concentration of receptor.  $K_D$  was determined by averaging of the  $K_D$  determined at the 3 concentrations of CarD-BODIPY, each conducted in triplicate.

$$\text{Equation (2)} \quad P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}}$$

$$\text{Equation (3)} \quad \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC_{50} - X) \cdot \text{Hill Slope}}}$$

### *Competition Experiments*

To determine  $IC_{50}$  and  $K_i$  for WT and mutant CarDs, competition experiments were performed as follows. In a 20  $\mu$ L final volume of 1x CarD FP Buffer 15 nM MTB RNAP (1.25x the  $K_D$ ) and 300 nM rrnAP3 DNA fragment were kept constant. For determination of  $K_i$ , CarD-D68C-BODIPY was tested at 1 nM, 3 nM, and 5nM final concentration. WT, R47E, and K90A mutant CarD each were each tested from 20.5  $\mu$ M down to 625 pM. For determination of small molecule  $IC_{50}$  competition experiments, CarD-BODIPY was held

constant at 3 nM and DMSO was added to a final concentration of 5% (v/v). IC<sub>50</sub> was determined using **Equation (4)** and K<sub>i</sub> was determined from **Equation (5)** from 3 independent experiments, each conducted in triplicate, at 3 CarD-BODIPY concentrations which were averaged for the final K<sub>i</sub>. Where [I]<sub>50</sub> is the concentration of competitor at 50% inhibition (i.e., IC<sub>50</sub>), [L]<sub>50</sub> is the concentration of free labeled ligand at 50% inhibition, [P]<sub>0</sub> is the concentration of free protein at 0% inhibition, and K<sub>D</sub> is the dissociation constant for the CarD•RPO complex.

$$\text{Equation (4)} \quad Y = \frac{100}{1+10^{((\log IC_{50}-X)*\text{Hill Slope})}}$$

$$\text{Equation (5)} \quad K_i = \frac{[I]_{50}}{\frac{[L]_{50}}{K_D} + \frac{[P]_{50}}{K_D+1}}$$

### *Pilot Screen*

A pilot screen of 237-pure small molecules from AnalytiCon was conducted as follows. Reaction conditions are described above with the following differences, 1 μL of each compound was deposited into individual wells followed by MTB RNAP, which was allowed to pre-incubate with compound for 20 minutes at 37°C prior to the addition of rrnAP3 DNA and CarD-D68C-BODIPY, final concentration of each compound was 25 μM. All components were allowed to incubate at 37°C for an additional 2 hours before reading the plate as mentioned above. For the negative control, reactions contained DMSO and for the positive control, RNAP was excluded. The z-score (Z) was calculated using **Equation (6)**, where σ<sub>P</sub> and σ<sub>N</sub> are the standard deviation of the positive and negative controls and μ<sub>P</sub> and μ<sub>N</sub> are the mean of the positive and negative controls respectively. Hits were characterized as those which caused a decrease in mP greater than 3 standard

deviations from the negative controls. Compounds which met these criteria were tested in triplicate. For compounds that reconfirmed, concentration-response curves were generated.

$$\text{Equation (6)} \quad Z = 1 - \frac{3(\sigma_P + \sigma_N)}{|\mu_P - \mu_N|}$$

### *Primary High Throughput Screen*

A high throughput screen (HTS) of the Maybridge 24K (MB24K) library (23,320 compounds) was conducted using the CarD FP assay. Assay conditions were similar to those for the pilot screen described above with the following differences. The CarD FP assay was conducted at 25°C, therefore the buffering conditions were altered so that the pH was consistent with conditions tested above at 37°C. A shift in  $K_D$  for the interaction of CarD with the RNAP•DNA complex (8 nM at 25°C, see **Appendix Figure IV-A10**) resulted in the use of 10 nM of RNAP (1.25x  $K_D$ ) for the screen. CarD-D68C-BODIPY was kept constant at 5 nM and rrnAP3 at a constant 300 nM. All other buffer conditions were as described above.

First, 10 µL of 2x MTB RNAP was added to the 384-well plate using a multidrop dispenser (Thermo Fisher Scientific) followed by 200 nL of compound or DMSO using a Biomek FX HDR pintool instrument (Beckman Coulter) so that final concentration in the assay was 20 µM. Compounds were allowed to incubate with MTB RNAP for 20 minutes before addition of 10 µL of 2x CarD-D68C-BODIPY and DNA, again with the multidrop dispenser to a final volume of 20.2 µL. For negative controls DMSO without compound was used and for positive controls MTB RNAP was excluded. Reactions were allowed to

come to equilibrium at 25°C for 2 hours, after which plates were read on a PHERAstar FSX plate reader using 485/20 excitation and 520/20 emission fluorescence polarization filters.

#### *Confirmation, Counter Screen, and Concentration Response Screen*

Compounds which met hit criteria (see **Figure IV-8**) were tested as in the primary screen in triplicate. A counter screen was conducted to detect compound interference with CarD-D68C-BODIPY fluorescence in the absence of MTB. This assay was conducted similar to that of the primary and confirmation screen with the following difference. Substituting for the 10  $\mu$ L 2x MTB RNAP, 10  $\mu$ L of 1x reaction buffer was added to the well with the multidrop dispenser. All compounds which were processed for confirmation were tested in this assay, again in triplicate. Plates were handled in the same fashion as for the primary screen. Compounds which confirmed and met criteria set for the counter screen were tested in duplicate for concentration-dependent inhibition and were derived from separated stock plates. For concentration-response studies, 200 nL of compound (final concentrations were varied from 150 – 4.2  $\mu$ M; 1.67x dilutions) or DMSO were dry spotted into 384-well plates prior to addition of any of the assay components using a Mosquito X1 dispenser and plates were centrifuged. Assay components were then added as described above.

#### *Reconfirmation*

Reconfirmations with purchased, fresh solid/powder compounds were performed essentially as described above. Fresh compounds were resuspended in DMSO and 1  $\mu$ L deposited into a 384-well plate (Final concentration: 1000  $\mu$ M to 0.46  $\mu$ M). To the 384-well

black low volume U-bottom plate 1  $\mu\text{L}$  of compound was added. Using the multidrop, 9  $\mu\text{L}$  of a 22.2 nM RNAP solution in 1x CarD FP buffer was added and allowed to incubate with the compounds for 20 minutes at 25°C before addition of 10  $\mu\text{L}$  of a 2x CarD D68C BODIPY and rrnAP3 DNA solution. The reaction was allowed to come equilibrium over 2 hours at 25°C and then read on the PHERAstar or Biotek H1 Synergy plate reader. The data were processed as previously described.

### *Surface Plasmon Resonance*

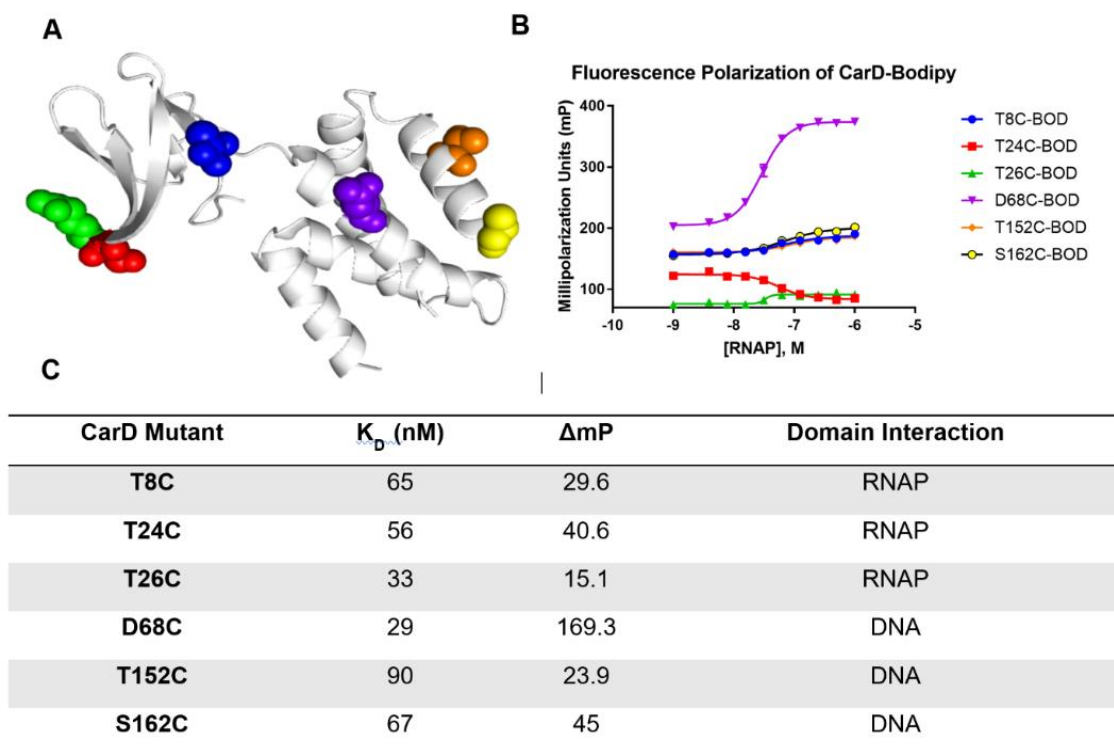
Surface plasmon resonance experiments were performed on a SensiQ Pioneer instrument using a HisCap biosensor (Pall FortéBio) with 3 channels in series. N-terminal decahistidine tagged CarD was prepared as described in Chapter III with the following modification: SEC buffer used during the final purification step contained 10 mM HEPES (pH 8.0), 300 mM NaCl, and 5% glycerol. CarD was stored as previously described. Channel 1 was charged 20  $\mu\text{L}$  of 50  $\mu\text{M}$   $\text{NiCl}_2$  at a flow rate of 10  $\mu\text{L}/\text{minute}$  followed by His-tagged CarD which was immobilized by injecting a 4  $\mu\text{g}/\text{mL}$  solution of CarD suspended in CarD SPR buffer (10 mM HEPES (pH 8.0 at 25°C), 150 KCl, 0.01% Triton X-100 and 5% DMSO) for 8.5 minutes at an injection speed 10  $\mu\text{L}/\text{min}$ . All compounds tested were lyophilized prior to being resuspended in SPR buffer. Compounds were tested at 100  $\mu\text{M}$ , 75  $\mu\text{M}$ , 50  $\mu\text{M}$ , 25  $\mu\text{M}$ , and 12.5  $\mu\text{M}$  (unless otherwise noted) in a random order with periodic blanks every 5 sample injections at a sampling rate of 10 Hz. Each injection consisted of a 100  $\mu\text{L}$  of sample at a flow rate of 50  $\mu\text{L}/\text{minute}$ . Response curves were normalized to blank injections and association rate ( $k_a$ ), dissociation rate ( $k_d$ ), and binding constant ( $K_D$ ) were determined by fitting to a 1:1 interaction model using Qdat software. As a counter screen, all compounds which showed concentration dependent

response with CarD were tested for binding to eIF4E, a human translation factor which was a gift from the Garner lab. Due to eIF4E's poor thermal stability, reactions with eIF4e were run at 15°C.

## Results

### *CarD Fluorescence Polarization Assay*

To develop the fluorescence polarization (FP) assay, site-specific mutagenesis of several selected residues on CarD to cysteine was performed followed by chemical modification by BODIPY iodoacetamide. Labeling efficiency of the CarD mutants varied from ~40%-100% (see **Appendix Table IV-A2**). Labeling sites were explored on both the N-terminal RNAP-interacting domain (RID) and the C-terminal DNA-interacting domain (DID). The dynamic range and  $K_D$  of each of the BODIPY-labeled CarD mutant were characterized by titration with WT MTB RNAP (see **Figure IV-1**).



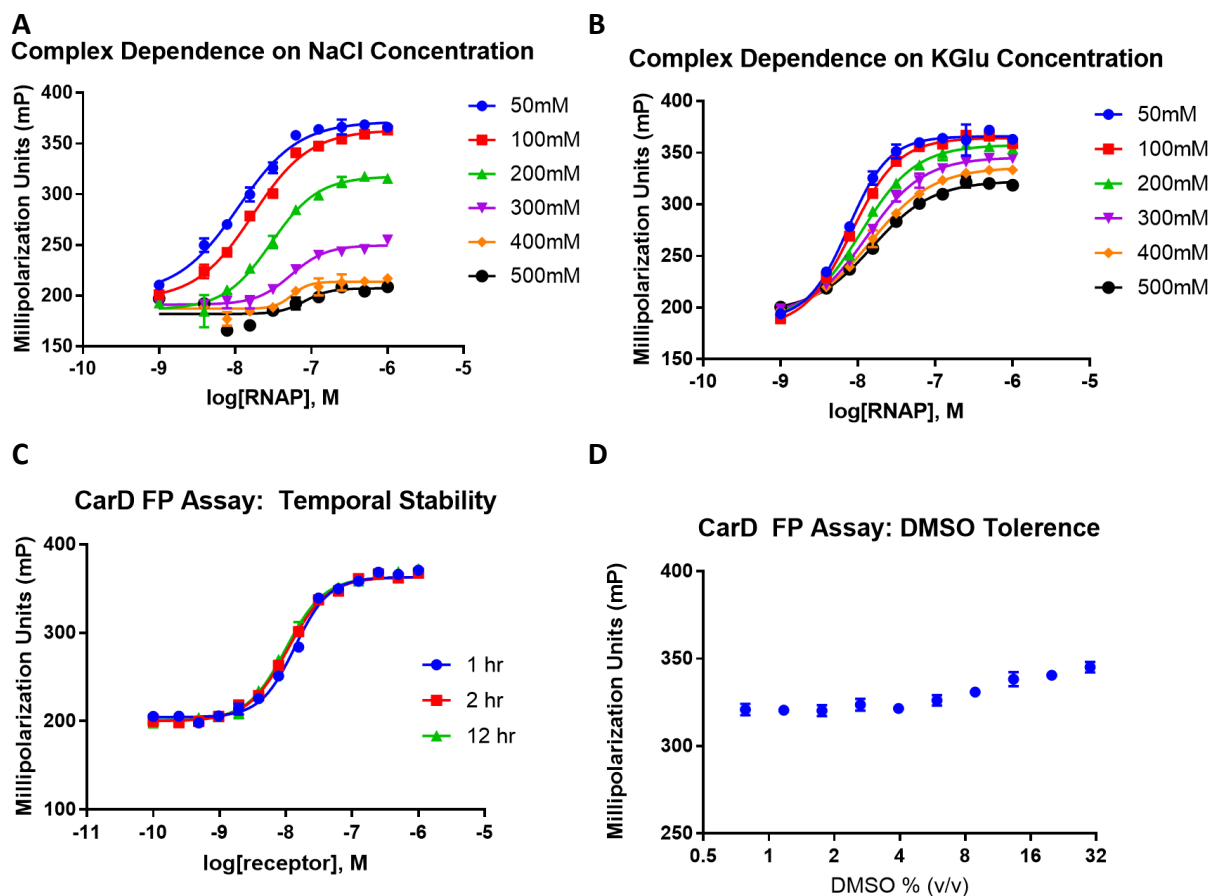
**Figure IV-1: Characterization of labeling sites on CarD.** **A)** Sites on CarD which were selected for site-specific mutagenesis and subsequent labeling with BODIPY FL (colors correspond with binding isotherms in **B**). **B)** Binding isotherms for CarD mutants labeled with BODIPY FL (note: several differences in this assay are as follows: 50 mM KGlu, 0.001% Triton X-100, and 12.5 nM BODIPY-labeled CarD were used in this experiment). **C)** Summary of isotherm parameters for each CarD variant as well as domain interaction on the RPo where BODIPY probe is located.

Several mutant labeled CarD's produced discernable binding isotherms, though the range and affinity varied for each of the mutants (see **Figure IV-1**). Of the six variants explored, labeling at position D68 produced a binding isotherm with the greatest dynamic range ( $\Delta mP = 169.3$ ) and had the strongest affinity for the RPo ( $K_D = 29$  nM, under the conditions tested). Several other labeled CarD mutants also produced binding isotherms (T8C and S162C) though the dynamic range was significantly smaller ( $\Delta mP \sim 40$ ) and they had lower affinity ( $K_D \sim 65$  nM), which may suggest the probe could be interfering

with binding. There were small responses,  $\Delta mP = 23.9$  and  $\Delta mP = 15.1$ , for CarD T152C and CarD T26C respectively. For CarD T26C an inverse binding isotherm was observed, possibly due to the propeller effect on the BODIPY probe upon binding to the RPo complex. The labeled D68C CarD mutant was used for all further CarD FP studies.

Optimal binding conditions for the CarD•RPo interaction were explored as was the stability of the FP in the presence of DMSO and over time (see **Figure IV-2**). Binding isotherms for CarD•RPo was monitored in titration experiments with both NaCl and potassium glutamate (KGlu) (see **Figure IV-2A-B**). Increasing concentrations of NaCl resulted in complex destabilization above 100 mM NaCl. The ternary complex was significantly more resistant in the presence of KGlu with isotherm contours still discernable up to 500 mM KGlu though a decrease in dynamic range is evident. Final salt conditions were a final concentration of 150 mM KGlu. The assay signal was also found to be stable over 24-hours (see **Figure IV-2C**). Addition of DMSO to ternary complex at RNAP concentrations (1.25x the  $K_D$  or 15 nM, see below) shows that the assay can tolerate DMSO up to 5% without significant deviation from the negative control (see **Figure IV-2D**).

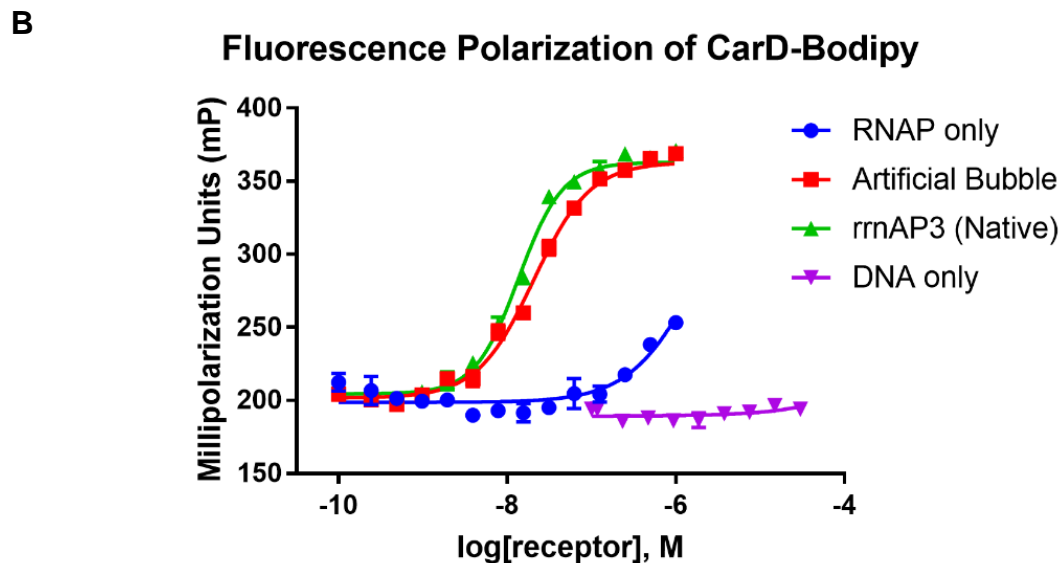
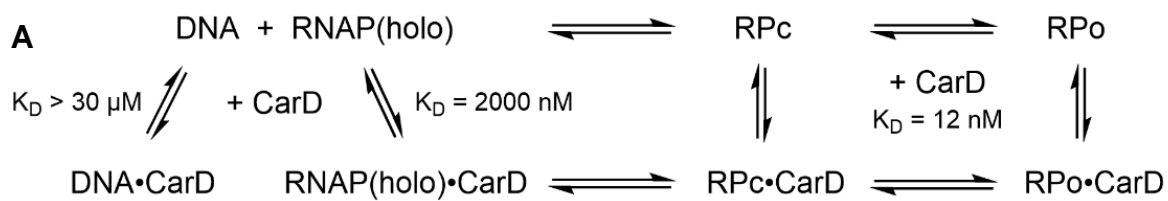




**Figure IV-2: Optimization of CarD FP Assay. A)** Stability of the CarD FP assay with NaCl titration. **B)** Stability of CarD FP assay with KGluc titration. **C)** Temporal stability of the CarD FP Assay at 37°C (for stability at 25°C see **Appendix Figure IV-A10**). **D)** Tolerance of CarD assay in the presence of increasing DMSO.

Binding isotherms were produced using labeled CarD with promoter DNA and RNAP only, as well as, with both native rrnAP3 RPo and a DNA fragment with an artificial bubble from -11 to +1 to determine dissociation constants,  $K_D$  (see **Figure IV-3**). Dissociation constants for each of the components mentioned above were determined in three independent experiments done in triplicate with varying concentrations of labeled CarD (1nM, 3nM, and 5nM), the average of all three experiments were considered for final calculations of  $K_D$ 's (see **Figure IV-3**). No polarization response was observed for

DNA fragments up to 30  $\mu\text{M}$ . RNAP only (no DNA) concentrations were tested up to 1  $\mu\text{M}$ ; however, a complete binding isotherm could not be generated from the concentrations tested. An estimate of  $\sim 2 \mu\text{M}$  for the RNAP•CarD  $K_D$  was made by extrapolating the RNAP only binding curve to the maximum mP observed in the presence of DNA. The affinity for the RPo using the “native” rrnAP3 DNA fragment was determined to be  $12 \pm 2 \text{ nM}$  and for the artificial bubble DNA  $18 \pm 2 \text{ nM}$  (see **Figure IV-3**).



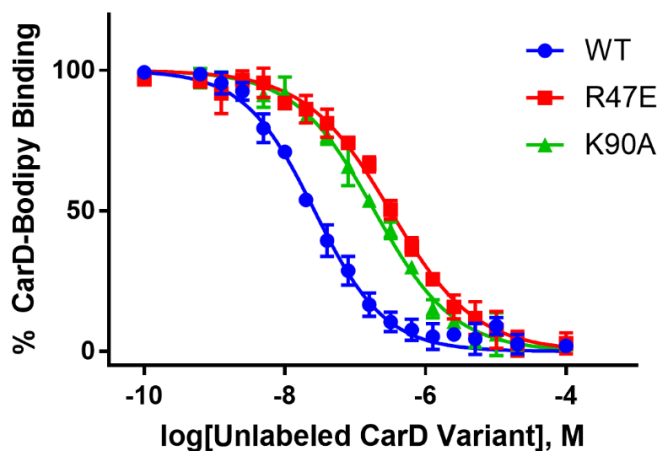
**Figure IV-3: CarD affinities for DNA, RNAP, and the open-promoter complex (RPo).**  
**A)** Schematic of CarD binding partners with dissociation constants derived from **B**. **B)** FP of CarD-Bodipy with assay components.

### *Competition and Preliminary Inhibition Experiments*

Competition experiments were performed using 15 nM RNAP (1.25x  $K_D$ ) and “native” rrnAP3 DNA template at 300 nM which was determined to be the minimal amount of DNA needed to maintain RPo complex stability (see **Appendix Figure IV-A11**). Experiments were performed in triplicate at three concentrations of labeled CarD (1 nM, 3 nM, and 5 nM). The inhibition constant ( $K_i$ ) for WT unlabeled CarD was determined to be  $8 \pm 1$  nM, which is similar to the  $K_D$  of CarD-D68C-BODIPY for the “native” RPo complex, indicating that the BODIPY probe does not interfere with CarD binding to RPo (see **Figure IV-4**). Inhibition constants were also determined for two CarD mutants, R47E which disrupts the CarD N-terminal interaction with RNAP and K90A which has been shown to disrupt CarD interactions with promoter DNA.<sup>9-10</sup> CarD R47E ( $K_i = 129 \pm 14$  nM) and CarD K90A ( $K_i = 85 \pm 6$  nM) show decreased affinity for the RPo complex, demonstrating these mutations are critical for CarD•RPo interaction (see **Figure IV-4**).

A

### Competition with Unlabeled CarD



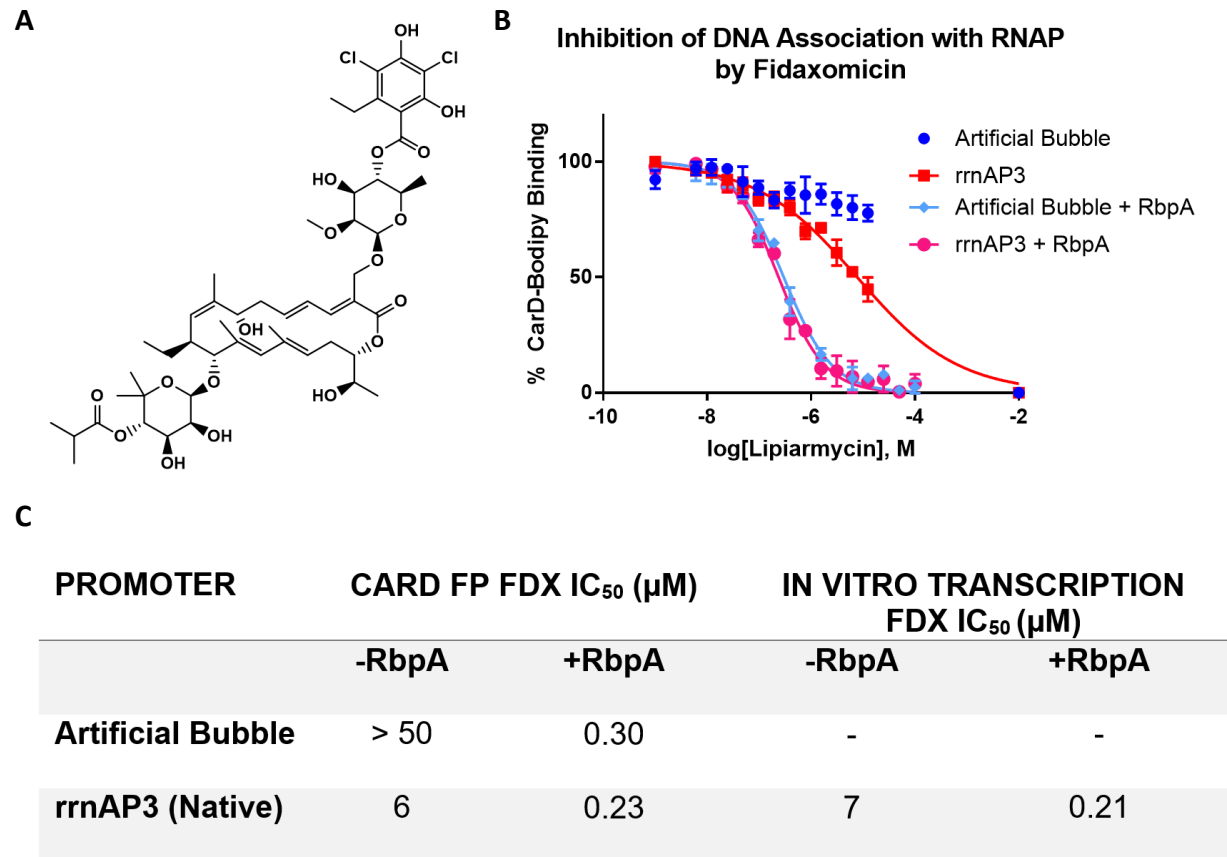
B

CarD Mutant	$K_D$ (nM)	$K_i$ (nM)	Interaction
WT	$12 \pm 2$	$8 \pm 1$	-
R47E	-	$129 \pm 14$	CarD-RNAP
K90A	-	$85 \pm 6$	CarD-DNA

**Figure IV-4: Competition experiments with unlabeled CarD variants. A)** Competition curve showing displacement of BODIPY-labeled CarD by WT and mutant CarD proteins (RNAP present at  $1.25 \times K_D$  as determined in **Figure IV-3**). **B)** Summary of binding and inhibition constants as determined using **Equation 5**.

To test the capability of the CarD FP assay to monitor the presence of DNA within RNAP, the  $IC_{50}$  of Fidaxomicin was determined (see **Figure IV-5**). Fidaxomicin (FDX) binds to the “switch” domain of RNAP locking it in an “open” clamp conformation preventing DNA from being securely bound within RNAP.<sup>11-13</sup> This experiment was conducted with both “native” and artificial bubble DNAs as well as in the presence and absence of *M. smegmatis* RNA polymerase binding protein, RbpA, which is required for Fidaxomicin potency against mycobacterial RNAPs. In the absence of RbpA, the observed  $IC_{50}$  is  $6 \mu\text{M}$  which is in agreement with inhibition observed with FDX in an *in vitro* transcription assay.<sup>12</sup> Interestingly, the  $IC_{50}$  determined with the artificial bubble DNA

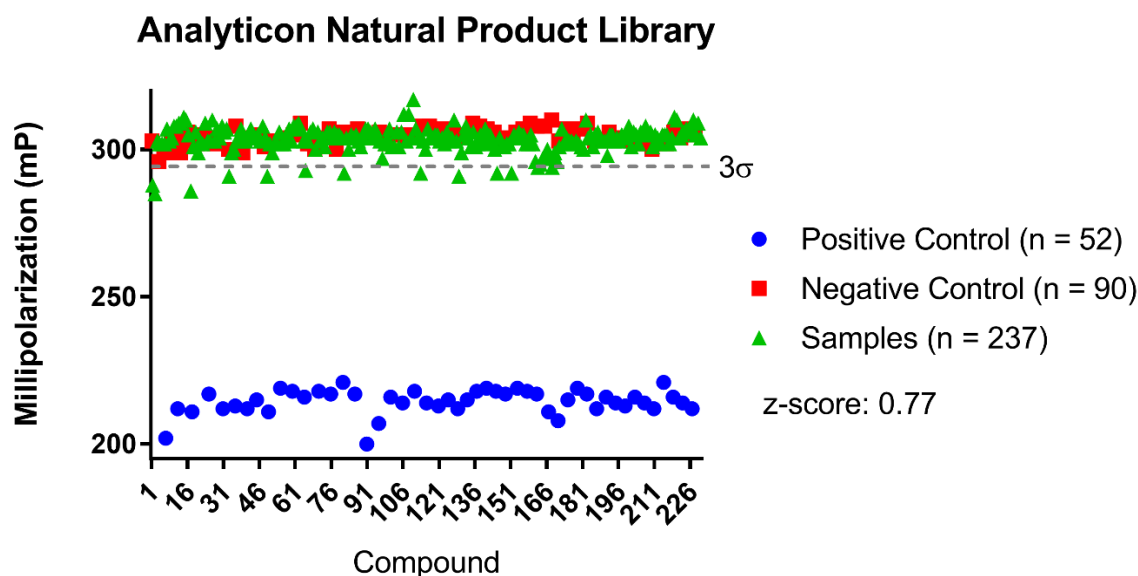
is significantly higher ( $IC_{50} < 80 \mu\text{M}$ ), the artificial bubble template shifts the equilibrium of RPo toward the “closed” clamp configuration which has a dramatically lower affinity for FDX. In the presence of RbpA, the “native” DNA template exhibits an increase in FDX potency to an  $IC_{50}$  of  $0.24 \mu\text{M}$ , which is consistent with other results recently reported.<sup>11</sup> In the presence of RbpA and the artificial bubble template, the  $IC_{50}$  for FDX is  $0.3 \mu\text{M}$ , which indicates that RbpA is able to overcome the reduction in affinity for FDX observed when using the artificial bubble template (see **Figure IV-5**).



**Figure IV-5: Fidaxomicin inhibition of CarD association with RPo.** **A)** Structure of Fidaxomicin. **B)** CarD binding to RPo in the presence of varying concentration of FDX with “native” rrnAP3 and artificial bubble DNA. **C)** Summary of inhibition data with and without RNA polymerase binding protein with both the CarD FP compared to literature values.

## Pilot Screen

A pilot screen of 237-pure small molecule library from AnalytiCon was conducted to evaluate the efficacy of the CarD FP assay for HTS (see **Figure IV-6**). For negative controls (n = 90) DMSO only was included (5% v/v) and for the positive control (n = 52) MTB RNAP was omitted. A Z-score = 0.77 was calculated using **Equation (6)** and the preliminary hit rate was determined to be 5% (12/237). All compounds were tested in triplicate and a single compound was found to confirm. It was determined that several false positives were identified from edge effects likely due to thermal variation along the edge of the plate (assay run at 37°C). During concentration response for the compound in question, it was determined that this compound likely has a fluorescent contaminant present in the sample and was also deemed a false positive. Though no hits were identified, this pilot demonstrated this assay is well suited for a larger HTS campaign.

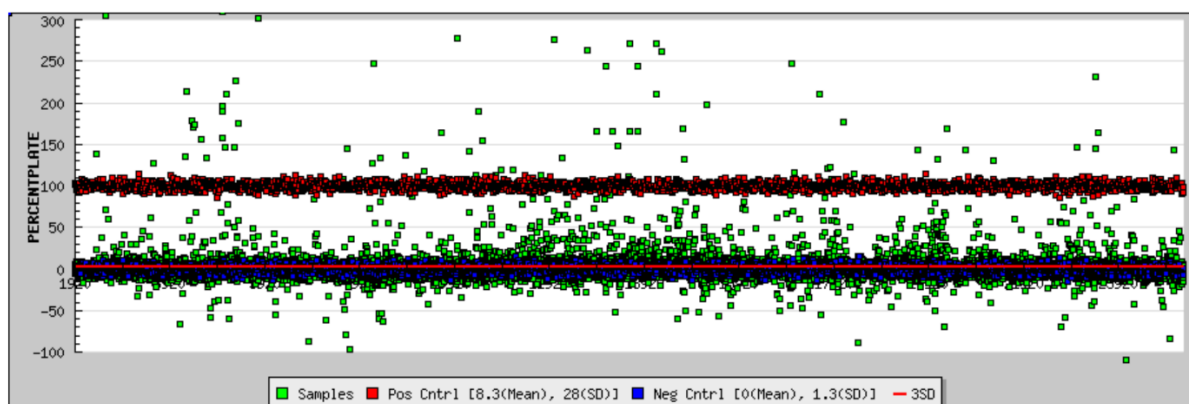


**Figure IV-6: Pilot screen run with CarD FP assay with AnalytiCon Library.**

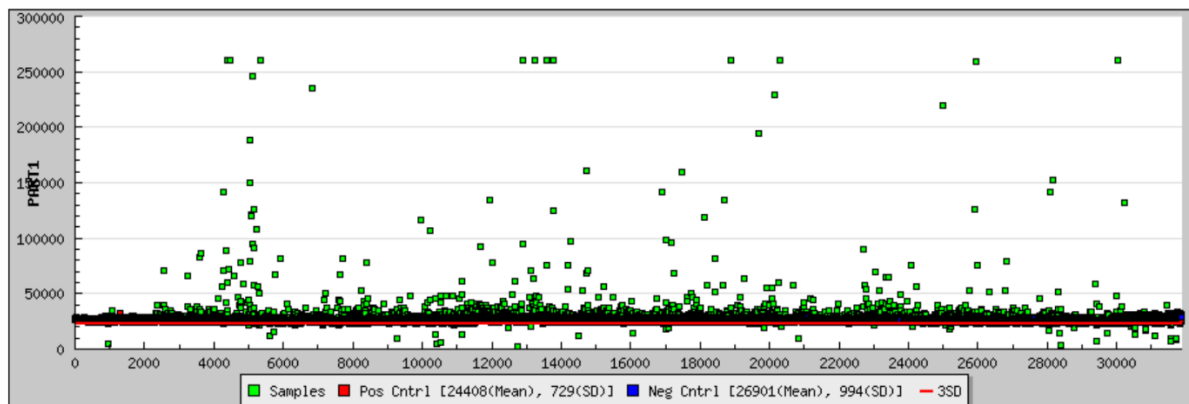
### Primary High Throughput Screen

A 23,320-compound small molecule library (Maybridge 24K) was screened using the CarD FP assay (see **Figure IV-7A**). Due to thermal edge effects observed in the pilot screen this assay was run at 25°C instead of 37°C. At this temperature the assay has a reduced dynamic range ( $\Delta mP = 138$ ) and a modest increase in the affinity of CarD for the RPo ( $K_D = 7.5$  nM) (see **Appendix Figure IV-A10**). For the primary screen 10 nM WT MTB RNAP (1.33x  $K_D$ ) was used which resulted in a dynamic range of  $\Delta mP = 76.9$ . The Z-score for the primary screen was determined to be 0.73.

A



B



**Figure IV-7: Maybridge 24K HTS with the CarD FP assay. A) Percent activity and B) parallel intensity of primary screen results (Positive controls (Red), negative controls (blue), and samples (green)).**

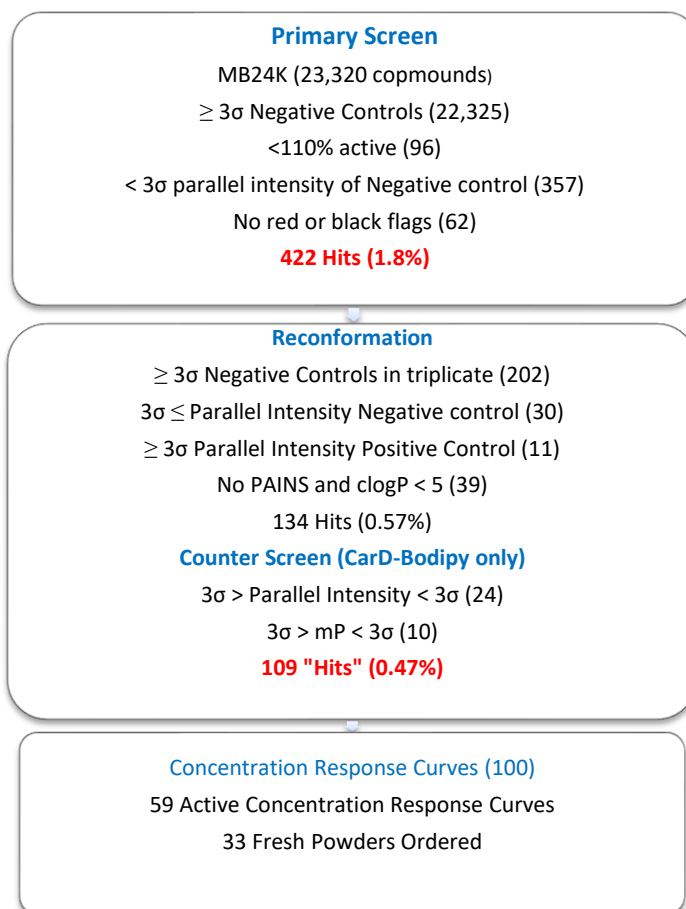
Several criteria were used to identify hits for this screen, 995 compounds or 4.26% were at least 3 standard deviations ( $3\sigma$ ) from the negative controls; however, there were a significant fraction of these which were found to be from compound interference with the assay and others which were red and black flagged by the NIH. Compounds which were >110% percent active (relative to the positive control on a plate by plate basis) were eliminated from the hit pool (96 compounds removed). These compounds are either false positives from fluorescence interference or are causing denaturation of the proteins in the assay. Additionally, compounds were eliminated from the hit pool if they caused an increase in the parallel intensity  $> 3\sigma$  from the average of the positive and negative controls or 29600 RFU (357 compounds removed, see **Figure IV-7B**). Compounds which were designated as red or black flag compounds were eliminated from the final count (62 compounds removed). Finally, a promiscuity filter was applied to eliminate compounds which hit in more than 9 other screens (58 compounds removed). After all filters were applied the final hit rate was 1.8% (422 / 23,320), which is substantially higher than the anticipated 1% hit rate.

#### *Confirmation, Counter Screen, and Concentration Response Screen*

All 422 compounds which were designated “hits” from the primary screen were re-tested in triplicate from the same stock plate of compounds. The Z-score for the retest plates was 0.76 (see **Appendix Figure IV-A12A**). Compounds which were  $>3\sigma$  from the negative controls in triplicate were considered confirmed hits (202 compounds removed). Additionally, compounds which were greater than or equal to  $3\sigma$  parallel intensity of the negative control and equal or less than  $3\sigma$  of the positive control were advanced (30 and 11 compounds removed respectively). All flagged PAINs (pan-assay interference



compounds) and compounds with a  $\text{clogP} > 5$  were removed (39 compounds removed total). At this point 134 compounds remained (0.57% of the primary screen). A counter screen was conducted to eliminate compounds which interfered with BODIPY-labeled CarD parallel fluorescence intensity (24 compounds removed) and fluorescence polarization (10 compounds removed) by more than  $3\sigma$  (see **Appendix Figure IV-A12B**). At this stage there were 109 compounds, or 0.47% of the primary screen remaining. Of the 109 compounds 100 were selected for concentration-response studies (CRC), 59 compounds were designated as a positive result in the CRC study (see **Figure IV-8**).



**Figure IV-8: Schematic of criteria for selection of hits from the primary screen to compound ordering (compounds removed with filter(s)).**

## Reconfirmation

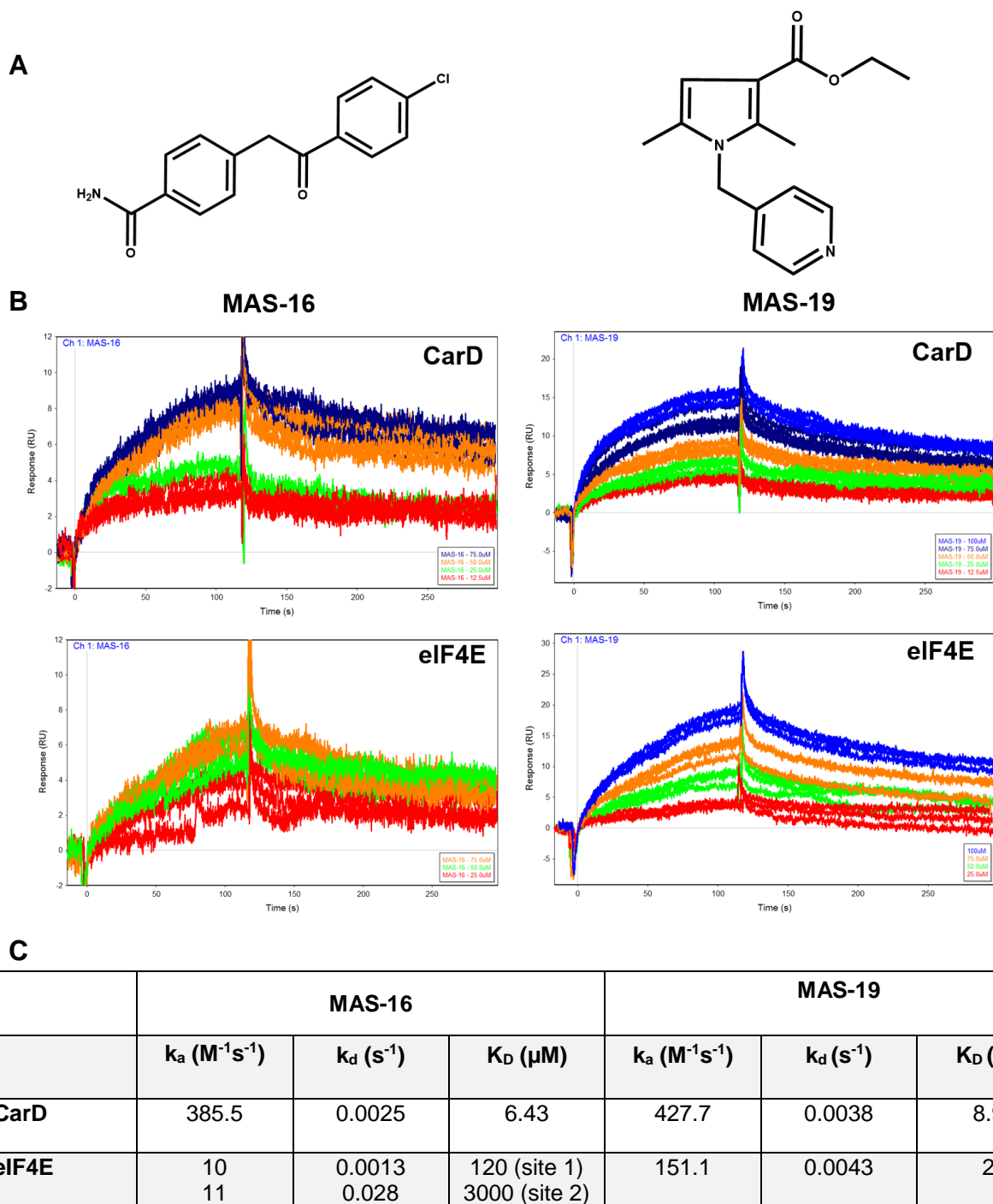
All 33 fresh powder samples were characterized by LCMS. Two of the 33 compounds were found to be either completely degraded or their MW did not match the labeled compound. All other fresh powders did match the MWs of the labeled compounds; however, the purities ranged from <50% to >95%, with the vast majority being the latter. Fresh powders ordered were tested in the CarD FP assay. Of the 33 compounds tested, 8 compounds were considered as hits (see **Table IV-1, Appendix Table IV-A3**), either because they have an IC<sub>50</sub> below 200  $\mu$ M, they are part of a cluster, or they showed direct binding via SPR (see below and **Appendix Figure IV-A13**). A single cluster (compounds with a similar structure) was identified (MAS-1 and MAS-2, see **Table IV-9**). Though very weak inhibitors, three of the compounds identified have IC<sub>50</sub>s less than 100  $\mu$ M. Eleven of the fresh powders did not reconfirm and several other compounds had extremely weak inhibition (>>200  $\mu$ M). One compound was found to be a non-specific binder by SPR. The physiochemical properties of these compounds can be found in **Appendix Table IV-A3**.

Name	Cluster	CCG CRC ( $\mu$ M)	Powder CRC ( $\mu$ M)
MAS-1	I	37	680
MAS-2	I	78	160
MAS-16	II	106	37
MAS-23	III	15	188
MAS-25	IV	250	132
MAS-26	V	90	47
MAS-28	VI	60	200
MAS-29	VII	106	85

**Table IV-1: Reconfirmed hits from fresh powders for the CarD MB24K HTS.**

## Surface Plasmon Resonance

All compounds ordered were screened by surface plasmon resonance (SPR) against decahistidine-tagged CarD which was immobilized on a HisCap biosensor. Of the 33 compounds tested (in duplicate at 75  $\mu\text{M}$ ), 11 compounds elicited a response. A full concentration dependent screen was then conducted with all 11 compounds. Of these 11 compounds 4 confirmed, 2 are designated as weak binders (MAS-1 and MAS-2) because of the overall low response ( $< 5$  response units (RU)) and 2 compounds are binding CarD in a concentration-dependent manner (MAS-16 and MAS-19). MAS-16 is the most promising of the set ( $K_D = 6.4 \pm 0.02 \mu\text{M}$ ), it has a slow on rate of  $385.5 \pm 0.8 \text{ M}^{-1}\text{s}^{-1}$  and slow off rate of  $2.48 \pm 0.004 \times 10^{-3} \text{ s}^{-1}$  (see **Figure IV-9**). Even though MAS-19 has a significantly weak  $\text{IC}_{50}$  it appears to have a  $K_D$  of  $8.91 \pm 0.02 \mu\text{M}$  with CarD. MAS-1, -2, -16, and -19 were also screened against eukaryotic initiation factor 4E (eIF4E), which is involved in translation in humans to determine whether the observed response with CarD was non-specific (CarD and eIF4E share 8.4% sequence similarity). MAS-19 was found to bind to eIF4E ( $K_D = 28 \pm 0.1 \mu\text{M}$ ) in a similar fashion to CarD so it was eliminated as a potential hit (see **Figure IV-9**). Response curves for MAS-16 for eIF4E had to be fit with a two-state binding fit due to the biphasic nature of the responses. MAS-16 showed very low binding to eIF4E, with an affinity of  $120 \pm 30 \mu\text{M}$  for the first site and  $3 \pm 1 \text{ mM}$  for the second site. Saturation was never reached for MAS-16 against eIF4E at any concentration of MAS-16 tested. Due to the limit of compound solubility, MAS-16 was only tested at 12.5  $\mu\text{M}$ , 25  $\mu\text{M}$ , 50  $\mu\text{M}$  and 75  $\mu\text{M}$  (see **Appendix Figure IV-A13**).



**Figure IV-9: Monitoring direct binding of small molecules to CarD and eIF4E by SPR. A)** Structures of MAS-16 and MAS-19. **B)** Response curves for both MAS-16 and MAS-19 for CarD and eIF4E. **C)** Summary of kinetic data derived from binding data in B.

## Discussion

Identification of new antibiotics which act by novel mechanisms of action is critical for addressing rising levels of antibiotic resistance. Targeting transcription in bacteria is a viable method for bacterial killing, as evidenced by the success of both Rifampin and Fidaxomicin. Traditionally, people have targeted RNA polymerase; however, productive transcription is a complex process which is highly regulated. We believe that this regulation represents an attractive target for novel antibiotic discovery. CarD is a global transcriptional regulator which has been shown to be required for MTB viability.<sup>3, 9, 14</sup> While the effect of CarD on other organism's viability has not been explored, its presence in other clinically relevant organisms represents a unique opportunity to be a novel broad-spectrum target.

A highly robust fluorescence polarization assay which probes the interaction between CarD and the RPo (open-promoter complex) was developed and utilized to identify novel antibiotic chemical matter. CarD lacks endogenous cysteine residues which allowed for site-specific incorporation of a BODIPY FL probe using site-directed mutagenesis and thiol reactive conjugated probe (see **Figure IV-1**). Percent labeling was monitored both spectroscopically and via LCMS. LCMS results indicate >90% labeling in all cases with the exception of CarD S162C mutant, where labeling was determined to be 40% by UV/Vis methods and 50% by LCMS (see **Appendix Table IV-A2**). This is likely because cysteine at this position has a pKa which is significantly higher than the standard pKa ~8.0 of cysteine due to the proximity to the C-terminus.<sup>15-16</sup> Differences in labeling efficiencies between the UV/vis methodology and LCMS are likely due to the local environment of the BODIPY probe causing a shift in the observed excitation or emission

maxima or could be effecting the extinction coefficient or quantum yeild;<sup>17-18</sup> therefore, we believe the LCMS results more accurately reflect percent labeling of CarD mutants.

Not surprisingly NaCl cause destabilization of the RPo which was exacerbated with increasing concentrations of NaCl (see **Figure IV-2A-B**).<sup>4</sup> The Cl<sup>-</sup> anion has been reported to cause destabilization of the RPo and our data reflects previous reports of CarD efficacy on stabilization of the RPo.<sup>4</sup> The RPo was more resistant to challenge by KGlu, which is actually the more prevalent anion (glutamate) found *in vivo* making this assay more reflective of physiological conditions.<sup>19-20</sup> In preparing for the use of this assay for HTS, we found that our signal is stable for over 12 hours at 37°C (longer times not tested), but under screening conditions at 25°C the signal stability extended up to 48 hours (see **Figure IV-2C, Appendix Figure IV-A10B**). DMSO cause increase in polarization of CarD•RPo at concentrations above 10% (v/v), likely effecting the organization of water which would affect CarD tumbling in solution but did not have an effect at the maximum DMSO concentration used, 5% (v/v)(see **Figure IV-2D**).<sup>21-22</sup>

CarD requires DNA to be present for high affinity interaction RNAP (12nM vs 2 μM without DNA at 37°C). This binding affinity is consistent with predicted binding affinity determined by rate constants determined using stop flow experiments. Their reported value is 6.7 nM at 25°C whereas we recorded the interaction to have a dissociation constant of 7.5 nM at 25°C (see **Appendix Figure IV-A10**).<sup>5</sup> The interaction with the RPo is roughly 170x's tighter than when DNA is not present. Endogenous concentrations of CarD in *M. smegmatis* are reported to be 1.6 μM as determined from quantitative immunoblotting, which would mean CarD may be partially bound to free RNAP *in vivo*. No significant hyperpolarization was observed in DNA titration experiments.<sup>5</sup> The DNA

fragments used in this experiment at ~30,000 kDa which may not be large enough to elicit a response in the assay, typically a 10-fold difference in binding partner is required for a defined binding isotherm.

Inhibition constants were performed to validate the FP assay and to determine whether the BODIPY probe may be interfering with CarD bind to the RPo. Competition of BODIPY-labeled CarD with WT unlabeled CarD resulted in a  $K_i = 8$  nM, which is similar to the  $K_D$  determined for CarD affinity for the RPo (see **Figure IV-4**). This suggests that the probe has little to no effect on CarD binding to the RPo. Using two unlabeled CarD variants, R47E which it at the interface between CarD and RNAP and K90A which has been shown in EMSA experiments to be involved in DNA binding, competitions experiments show that the interaction between CarD and either the interaction with RNAP or DNA are compromised with these mutations. This suggests that if either of these functions are disrupted you can reduce the affinity of CarD and thus destabilize the RPo.

The differential binding affinity of CarD for free RNAP (~ 2  $\mu$ M) compared to its affinity for the RPo (12 nM) broadens the repertoire of types of inhibitors that can be identified with this assay (see **Figure IV-3**). Since we are conducting the HTS at 10 nM where there is no polarization response from free RNAP we should be able to identify inhibitors of DNA association with RNAP. This was shown to be the case with Fidaxomicin (FDX). FDX block RNAP clamp closure onto DNA preventing unwinding of DNA.<sup>11, 13</sup> Recent reports demonstrated the involvement of another transcriptional regulator RNA polymerase binding protein (RbpA) in high affinity inhibition of RNAP ( $IC_{50} = 0.21$   $\mu$ M).<sup>11</sup> RbpA makes contacts with its N-terminal tail (NTT) with FDX. Using *M. smegmatis* RbpA (which shares >90% sequence similarity to the MTB homolog) we show that we can

monitor the effect of FDX in preventing DNA binding in our CarD FP assay ( $IC_{50} = 0.24 \mu\text{M}$ ) (see **Figure IV-5**). Therefore, we believe this vastly expands the utility of this assay in drug discovery related to inhibition of transcription in MTB.

With the validated CarD FP assay a pilot screen on 237-purified small molecule natural products was conducted (see **Figure IV-6**). The Z-score for this assay was 0.77 which indicates this assay is highly robust and amenable for larger HTS campaigns. Though no inhibitors were discovered using this library the assay provided empirical support for running the larger screen. Additionally, thermal edge effects observed with the pilot screen prompted running the larger campaign at 25°C.

A 23,320-compound library (MB24K) was screening using the CarD FP assay. The z-score was 0.73 demonstrating the CarD FP assay has an excellent assay to use in a HTS format (see **Figure IV-7A**). The preliminary hit rate, based solely samples which were  $>3\sigma$  from the negative control, was 4.26%. This is significantly higher than the desired 1% hit rate for screening. BODIPY FL is a green-shift fluorophore, which can have significant overlapping excitation and emission spectra with compounds in libraries.<sup>23-24</sup> This is evident when you look at the parallel fluorescence intensity of the compounds in the primary screen (see **Figure IV-7B**). There are several reasons which may cause false positives such as autofluorescence and quenching by small molecules. To mitigate this, we eliminated compounds with parallel fluorescence intensities  $>3\sigma$  of the negative controls. Additionally, compounds which appeared to have  $>110\%$  activity in the assay were eliminated, several of these were likely due to compound interference, but may also be from compounds causing denaturation of protein and/or may be directly interfering with the BODIPY probe. None of these compounds would be useful for drug discovery.



Compounds which are red or black flags by the NIH were also removed, using these criteria we were able to reduce our hit rate to 1.8%, which was still high.

In reconfirmation 220 out of 422 compounds reconfirmed, reducing the hit rate to 0.94%, which is ideal for a screen of this magnitude. Compounds whose parallel fluorescence intensity which deviated from both the positive and negative control were removed (41 total). All PAINs and compounds which had a  $clogP > 5$  were also removed since these compounds would likely not be amenable lead candidates (39 compounds total), leaving us with 140 compounds or 0.57% of the primary screen. A counter screen was also run to eliminate compounds which altered parallel fluorescence intensity or mP of BODIPY-labeled CarD on its own. These compounds are likely directly effecting the BODIPY probe, removal of these compounds left us with 109 compounds or 0.47% of the primary screen (see **Figure IV-8; Appendix Figure IV-A12B,C**).

Concentration response curves (CRC) were generated for 100 of 109 initial hit compounds. Several compounds which appeared to be DNA intercalators, had reactive or unstable moieties (such as redox cyclers), or were chemically unattractive were removed prior to CRC analysis. Out of 100 compounds 59 demonstrated some extent of concentration response. We did not identify any compounds with an  $AC_{50} < 10 \mu M$ , which, while unfortunate, was not unexpected since we screened against a limited library and perhaps more importantly because we are trying to identify a protein-protein or protein-nucleic acid inhibitor. These types of inhibitors are notoriously difficult to identify because protein surfaces tend to lack contour and protein-nucleic acid interaction are usually, but not always non-specific.

Of the 33 fresh powder samples, only 8 are classified as hits at this point. Several compounds were eliminated either because of low potency, because they had steep hill slope (likely indicating inhibition due to aggregation), or they were non-specifically binding via SPR.<sup>25</sup> We set our upper limit at 200  $\mu\text{M}$  for several reasons: 1) we ran a limited screen, 2) protein-protein interactions are notoriously difficult to target, and 3) it is early on in the discovery process and we don't want to potentially eliminate false negatives.<sup>26</sup> A single cluster consisting of MAS-1 and MAS-2 was identified (see **Appendix Table IV-A3**).

All 33 compounds were initially screened in duplicate at 75  $\mu\text{M}$  against CarD via SPR. Of the 33, 11 elicited some response, though most were very weak binders. When we cross referenced these with the positives in our CarD FP assay, we had 4 compounds (MAS-1, -2, -16, and 1-9). MAS-1 and MAS-2 are very weak binders with responses below 5 RU; additionally, a concentration dependent response could not be readily observed, likely due to non-specific interactions. MAS-16 and MAS-19 gave maximum responses above 10 RU in a concentration dependent manner. As a counter screen, eIF4E was immobilized on the chip and both MAS-16 and MAS-19 were screened. A full concentration response from MAS-19 was also observed against eIF4E so this compound was eliminated from our hit list (see **Figure IV-9**). MAS-16 appears to have a significantly lower affinity for eIF4E ( $K_D = 120 \mu\text{M}$ ) compared to CarD ( $K_D = 6.43 \mu\text{M}$ ), which indicates this compound has selectivity for CarD. While these results are promising we are cautiously optimistic, as these are preliminary results. MAS-16 and the other compounds will need to be validated with several alternate secondary assays as well as show reproducibility with purified and validated compounds.

Future directions for this project include identifying potential binding sites for MAS-16 and MAS-19, in addition to determining whether the other identified hits are acting by preventing DNA association with RNAP. For the former project I propose using a computational approach by modeling in MAS-16 and MAS-19 to previously resolved structures of CarD using MOE (PDB: 4KBM). Models will be validated by using both alanine scanning and SPR to determine whether these residues are interacting with the compounds of interest. If these sites are validated, commercially available analogs should be purchased to improve potency. As for determining whether the other preliminary hits are disrupting the interaction between RNAP and DNA, the malachite green assay (described in **Chapter II**) or an electrophoretic mobility shift assay could be employed. While these are a few directions which could be explored, there are several directions which could be explored, including utilization of crystallography and NMR methodologies.

#### **Notes to Chapter IV**

I would like to thank the University of Michigan Center for Chemical Genomics their advice and supervision before, during, and after screening, especially Nick Santoro, Steve Roest, and Renju Jacob. I would also like to thank Dr. Amanda Garner for allowing us to screen her 237-compound natural product library. I would like to thank Dr. Katsuhiko Murakami and Catherine Sutherland from Penn State University for the *M. smegmatis* RbpA expression vector. I would especially like to thank Erin Gallagher for helping me design, execute, and interpret the data for the SPR experiments. I would also like to thank Jennifer Schmidt for help with the LCMS characterization of my CarD mutants and Tony Nastase, Brandt Huddle, and Jeff Zwicker for advice during the triage portion of the HTS.

## References

1. Weiss, L. A.; Harrison, P. G.; Nickels, B. E.; Glickman, M. S.; Campbell, E. A.; Darst, S. A.; Stallings, C. L., Interaction of CarD with RNA polymerase mediates *Mycobacterium tuberculosis* viability, rifampin resistance, and pathogenesis. *J Bacteriol* **2012**, *194* (20), 5621-31.
2. Hubin, E. A.; Fay, A.; Xu, C.; Bean, J. M.; Saecker, R. M.; Glickman, M. S.; Darst, S. A.; Campbell, E. A., Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife* **2017**, *6*.
3. Stallings, C. L.; Stephanou, N. C.; Chu, L.; Hochschild, A.; Nickels, B. E.; Glickman, M. S., CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell* **2009**, *138* (1), 146-59.
4. Davis, E.; Chen, J.; Leon, K.; Darst, S. A.; Campbell, E. A., Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* **2015**, *43* (1), 433-45.
5. Rammohan, J.; Ruiz Manzano, A.; Garner, A. L.; Stallings, C. L.; Galburt, E. A., CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Res* **2015**, *43* (6), 3272-85.
6. Bae, B.; Chen, J.; Davis, E.; Leon, K.; Darst, S. A.; Campbell, E. A., CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* **2015**, *4*.
7. Srivastava, D. B.; Leon, K.; Osmundson, J.; Garner, A. L.; Weiss, L. A.; Westblade, L. F.; Glickman, M. S.; Landick, R.; Darst, S. A.; Stallings, C. L.; Campbell, E. A., Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* **2013**, *110* (31), 12619-24.
8. Dey, A.; Verma, A. K.; Chatterji, D., Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA. *Microbiology* **2011**, *157* (Pt 7), 2056-71.

9. Garner, A. L.; Weiss, L. A.; Manzano, A. R.; Galburt, E. A.; Stallings, C. L., CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. *Mol Microbiol* **2014**, *93* (4), 682-97.
10. Gulten, G.; Sacchettini, J. C., Structure of the Mtb CarD/RNAP beta-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure* **2013**, *21* (10), 1859-69.
11. Boyaci, H.; Chen, J.; Lilic, M.; Palka, M.; Mooney, R. A.; Landick, R.; Darst, S. A.; Campbell, E. A., Fidaxomicin jams Mycobacterium tuberculosis RNA polymerase motions needed for initiation via RbpA contacts. *Elife* **2018**, *7*.
12. Buurman, E. T.; Foulk, M. A.; Gao, N.; Laganas, V. A.; McKinney, D. C.; Moustakas, D. T.; Rose, J. A.; Shapiro, A. B.; Fleming, P. R., Novel rapidly diversifiable antimicrobial RNA polymerase switch region inhibitors with confirmed mode of action in Haemophilus influenzae. *J Bacteriol* **2012**, *194* (20), 5504-12.
13. Lin, W.; Das, K.; Degen, D.; Mazumder, A.; Duchi, D.; Wang, D.; Ebright, Y. W.; Ebright, R. Y.; Sineva, E.; Gigliotti, M.; Srivastava, A.; Mandal, S.; Jiang, Y.; Liu, Y.; Yin, R.; Zhang, Z.; Eng, E. T.; Thomas, D.; Donadio, S.; Zhang, H.; Zhang, C.; Kapanidis, A. N.; Ebright, R. H., Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Mol Cell* **2018**, *70* (1), 60-71 e15.
14. Stallings, C. L.; Glickman, M. S., CarD: a new RNA polymerase modulator in mycobacteria. *Transcription* **2011**, *2* (1), 15-8.
15. Marino, S. M.; Gladyshev, V. N., Analysis and functional prediction of reactive cysteine residues. *J Biol Chem* **2012**, *287* (7), 4419-25.
16. Vollmar, B. S.; Wei, B.; Ohri, R.; Zhou, J.; He, J.; Yu, S. F.; Leipold, D.; Cosino, E.; Yee, S.; Fourie-O'Donohue, A.; Li, G.; Phillips, G. L.; Kozak, K. R.; Kamath, A.; Xu, K.; Lee, G.; Lazar, G. A.; Erickson, H. K., Attachment Site Cysteine Thiol pKa Is a Key Driver for Site-Dependent Stability of THIOMAB Antibody-Drug Conjugates. *Bioconjug Chem* **2017**, *28* (10), 2538-2548.
17. Toseland, C. P., Fluorescent labeling and modification of proteins. *J Chem Biol* **2013**, *6* (3), 85-95.

18. Albrecht, M.; Lippach, A.; Exner, M. P.; Jerbi, J.; Springborg, M.; Budisa, N.; Wenz, G., Site-specific conjugation of 8-ethynyl-BODIPY to a protein by [2 + 3] cycloaddition. *Org Biomol Chem* **2015**, *13* (24), 6728-36.
19. O'Hare, H. M.; Duran, R.; Cervenansky, C.; Bellinzoni, M.; Wehenkel, A. M.; Pritsch, O.; Obal, G.; Baumgartner, J.; Vialaret, J.; Johnsson, K.; Alzari, P. M., Regulation of glutamate metabolism by protein kinases in mycobacteria. *Mol Microbiol* **2008**, *70* (6), 1408-23.
20. Tian, J.; Bryk, R.; Itoh, M.; Suematsu, M.; Nathan, C., Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of alpha-ketoglutarate decarboxylase. *Proc Natl Acad Sci U S A* **2005**, *102* (30), 10670-5.
21. Daschakraborty, S., How do glycerol and dimethyl sulphoxide affect local tetrahedral structure of water around a nonpolar solute at low temperature? Importance of preferential interaction. *J Chem Phys* **2018**, *148* (13), 134501.
22. Idrissi, A.; Marekha, B. A.; Barj, M.; Miannay, F. A.; Takamuku, T.; Raptis, V.; Samios, J.; Jedlovszky, P., Local structure of dilute aqueous DMSO solutions, as seen from molecular dynamics simulations. *J Chem Phys* **2017**, *146* (23), 234507.
23. Hall, M. D.; Yasgar, A.; Peryea, T.; Braisted, J. C.; Jadhav, A.; Simeonov, A.; Coussens, N. P., Fluorescence polarization assays in high-throughput screening and drug discovery: a review. *Methods Appl Fluoresc* **2016**, *4* (2), 022001.
24. Simeonov, A.; Jadhav, A.; Thomas, C. J.; Wang, Y.; Huang, R.; Southall, N. T.; Shinn, P.; Smith, J.; Austin, C. P.; Auld, D. S.; Inglese, J., Fluorescence spectroscopic profiling of compound libraries. *J Med Chem* **2008**, *51* (8), 2363-71.
25. Jadhav, A.; Ferreira, R. S.; Klumpp, C.; Mott, B. T.; Austin, C. P.; Inglese, J.; Thomas, C. J.; Maloney, D. J.; Shoichet, B. K.; Simeonov, A., Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *J Med Chem* **2010**, *53* (1), 37-51.
26. Chene, P., Drugs targeting protein-protein interactions. *ChemMedChem* **2006**, *1* (4), 400-11.

## Appendix to Chapter IV

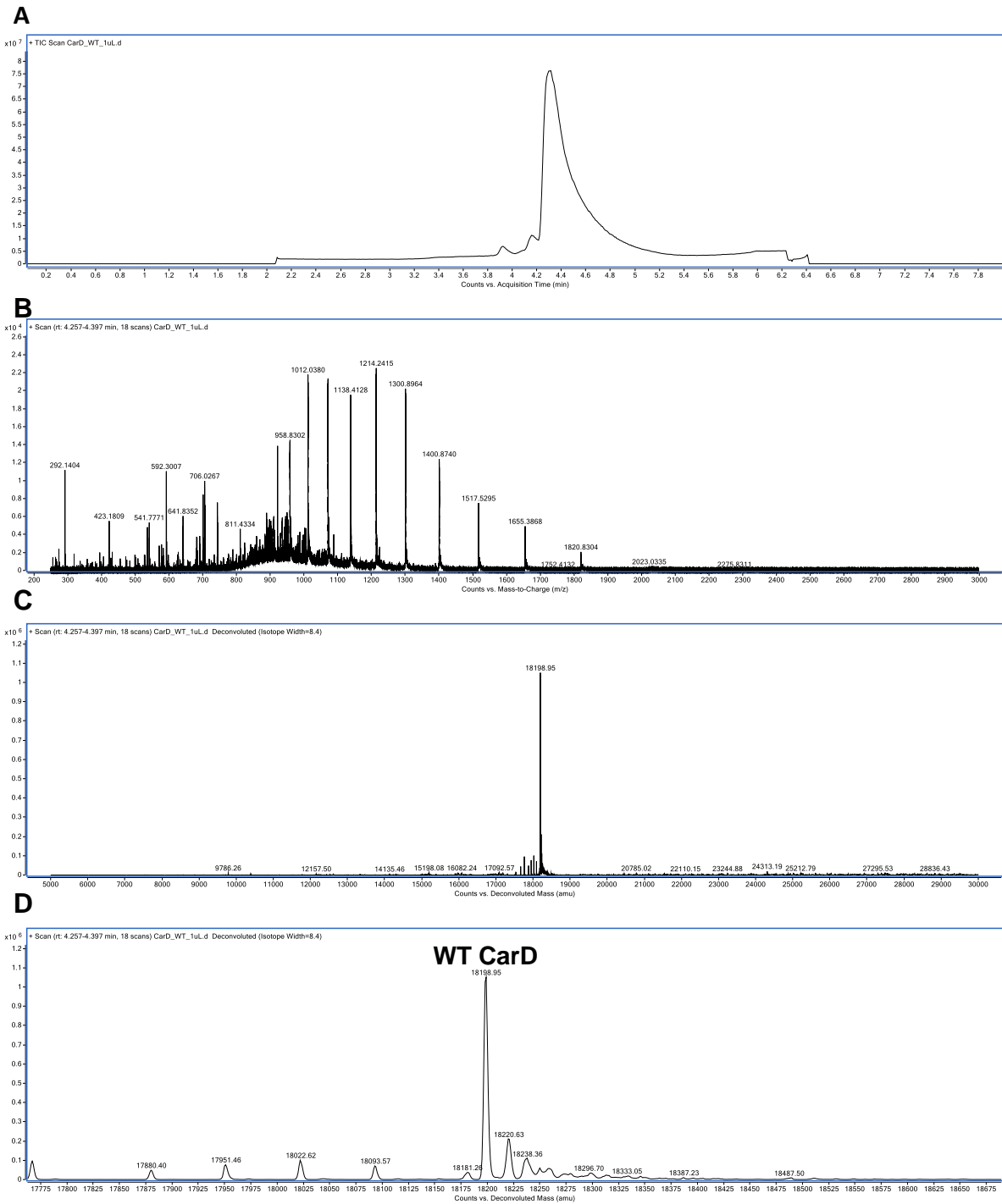
<b>Name</b>	<b>Sequence</b>	<b>Purpose</b>
<b>CarD T8C FOR</b>	ATTTTCAAGGTCGGAGACT <u>TC</u> CGTTGTCTATCCACACCAC	CarD T8C Mutagenesis
<b>CarD T8C REV</b>	GTGGTGTGGATAGACAACGC <u>AG</u> TCTCCGACCTTGAAAAT	CarD T8C Mutagenesis
<b>CarD T24C FOR</b>	GTCGAGGCGATCGAAT <u>GCC</u> GGACCATCAAAGGGG	CarD T24C Mutagenesis
<b>CarD T24C REV</b>	CCCCTTTGATGGTCCGG <u>C</u> ATTCGATCGCCTCGAC	CarD T24C Mutagenesis
<b>CarD T26C FOR</b>	CGATCGAAACCCGGT <u>GC</u> CATCAAAGGGGAAC	CarD T26C Mutagenesis
<b>CarD T26C REV</b>	GTTCCCCTTTGATG <u>C</u> ACCGGGTTTCGATCG	CarD T26C Mutagenesis
<b>CarD D68C FOR</b>	CGGGCAGGAAGGCCTGT <u>G</u> CAAGGTTTTCCAGG	CarD D68C Mutagenesis
<b>CarD D68C REV</b>	CCTGGAAAACCTTG <u>C</u> ACAGGCCTTCCTGCCCC	CarD D68C Mutagenesis
<b>CarD T152C FOR</b>	GACGCCAAAGCCGAGT <u>G</u> CATCCTTGACGAGG	CarD T152C Mutagenesis
<b>CarD T152C REV</b>	CCTCGTCAAGGATG <u>C</u> ACTCGGCTTTGGCGTC	CarD T152C Mutagenesis
<b>CarD S162C FOR</b>	GCCGCCGCGT <u>G</u> TTGAGGATCCGGCTGCTAACAAAGC	CarD S162C Mutagenesis
<b>CarD S162C REV</b>	GCTTTGTTAGCAGCCGGATCCTCAA <u>C</u> ACGCGGCGGC	CarD S162C Mutagenesis
<b>rrnAP3 FP TOP</b>	GATGACCGAACCTGGTCTTGACTCCATTGCCGGATTTGTATTAGACTGGCAGGGTT GCCCC	FP Assay DNA
<b>rrnAP3 FP BOTTOM</b>	GGGGCAACCCTGCCAGTCTAATACAAATCCGGCAATGGAGTCAAGACCAGGTTTCG GTCATC	FP Assay DNA
<b>Artificial Bubble TOP</b>	GGCTCTTGACAAAAGTGTTAAATTGTGCTATA <u>CTGGGATGG</u> TATGGATGACAGAAT TCGG	FP Assay DNA
<b>Artificial Bubble BOTTOM</b>	CCGAATTCTGTCATCCATAGGTAGGGTCATAAGCACAATTTAACACTTTTGTCAAGA GCC	FP Assay DNA

**Appendix Table IV-A1: Primers used in Chapter IV. Mutagenesis sites are underlined**

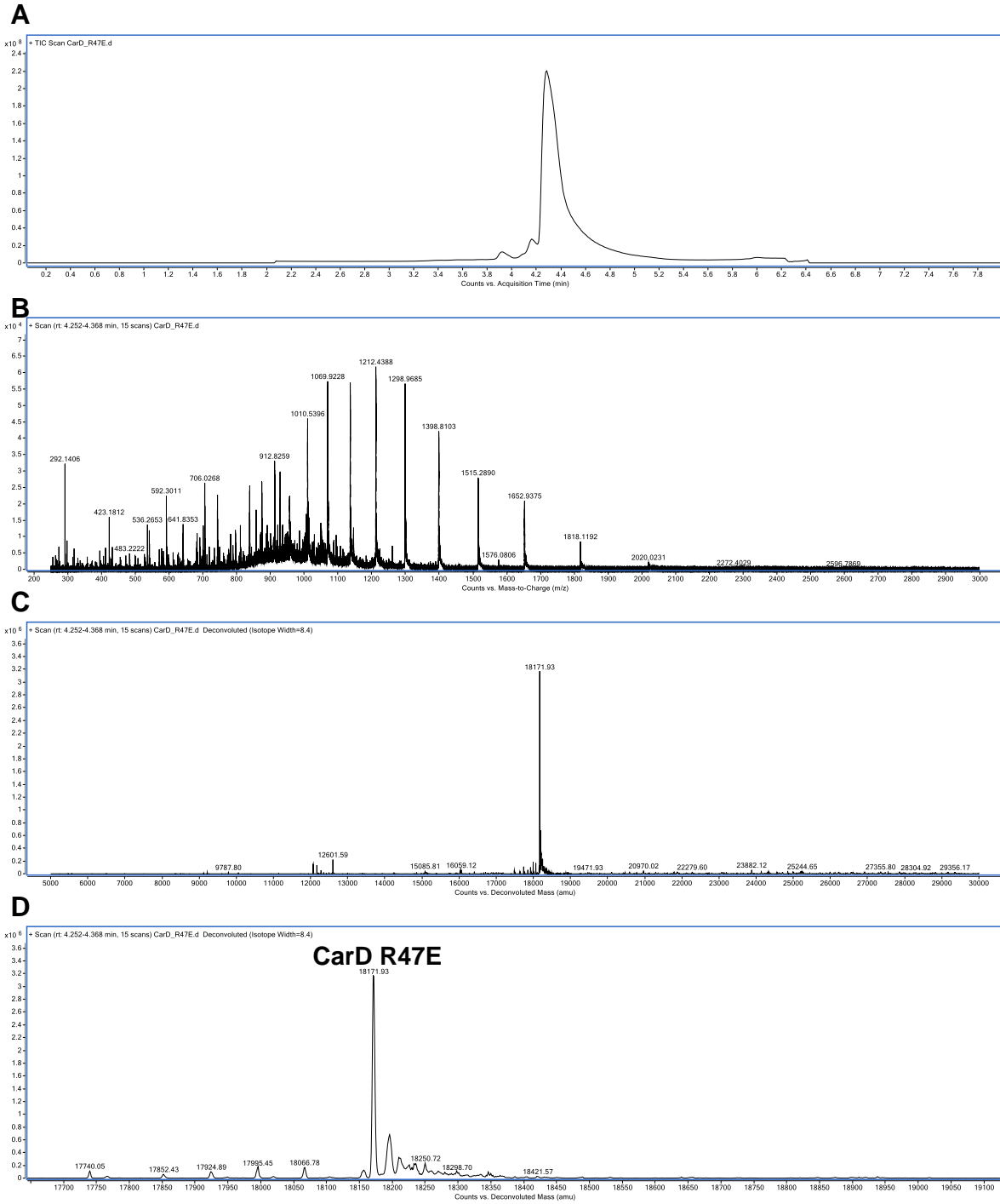
CarD Mutant	m/z (Calculated)		Percent Labeling	
	Unlabeled	Labeled	LCMS	UV/Vis
WT	18198.68	NA	NA	NA
R47E	18171.60	NA	NA	NA
K90A	18141.58	NA	NA	NA
T8C-BODIPY	18200.71	18489.81	91%	96%
T24C-BODIPY	18200.71	18489.81	97%	90%
T26C-BODIPY	18200.71	18489.81	98%	67%
D68C-BODIPY	18186.73	18475.83	100%	88%
T152C-BODIPY	18200.71	18489.81	98%	87%
S162C-BODIPY	18214.75	18503.85	52%	37%

**Appendix Table IV-A2: Characterization of mutant and BODIPY-labeled CarD variants.** Mass of CarD variants derived from theoretical values, refer to **Appendix Figures IV-1-8** for empirical values (all theoretical values are in agreement with empirical values).

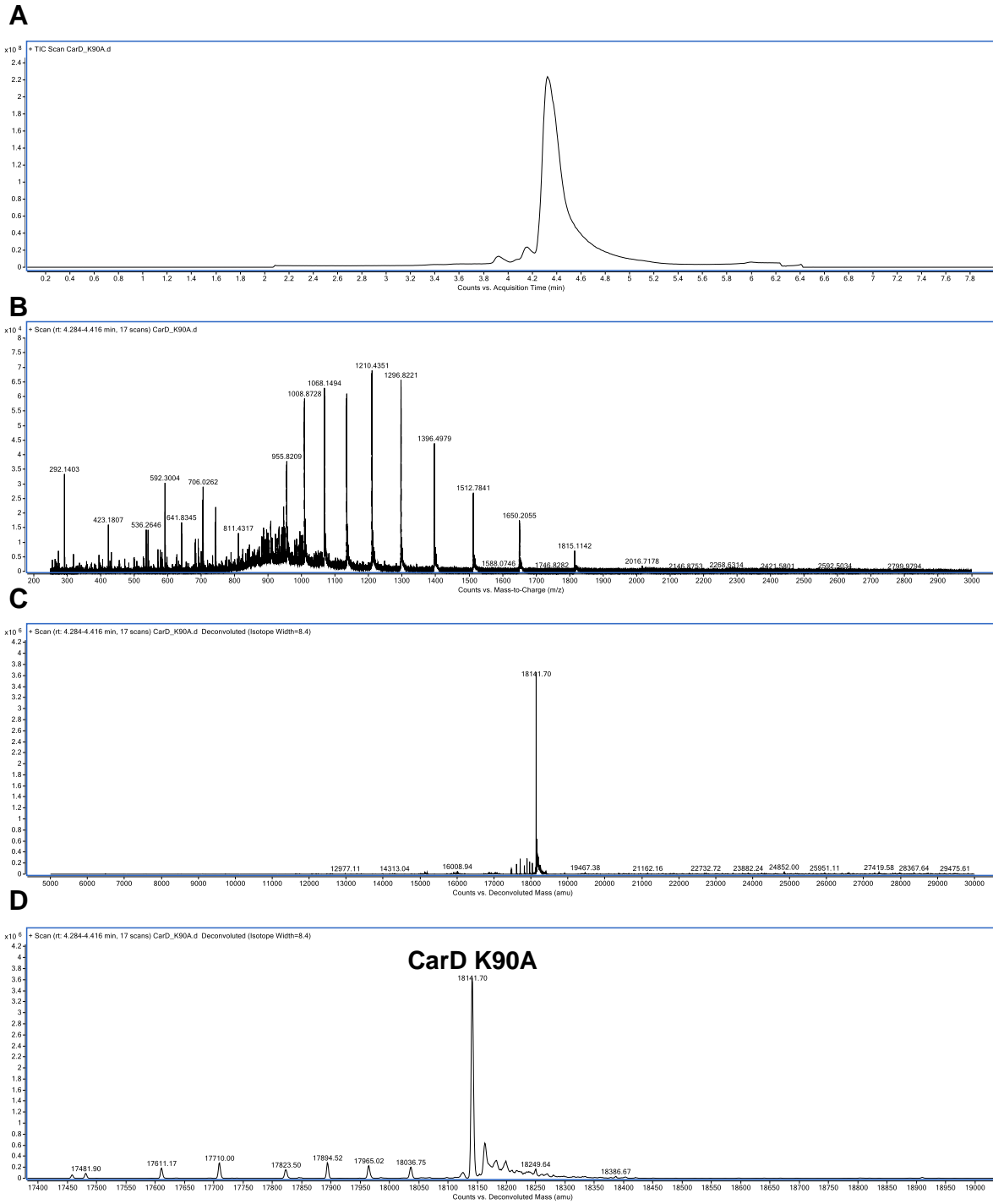




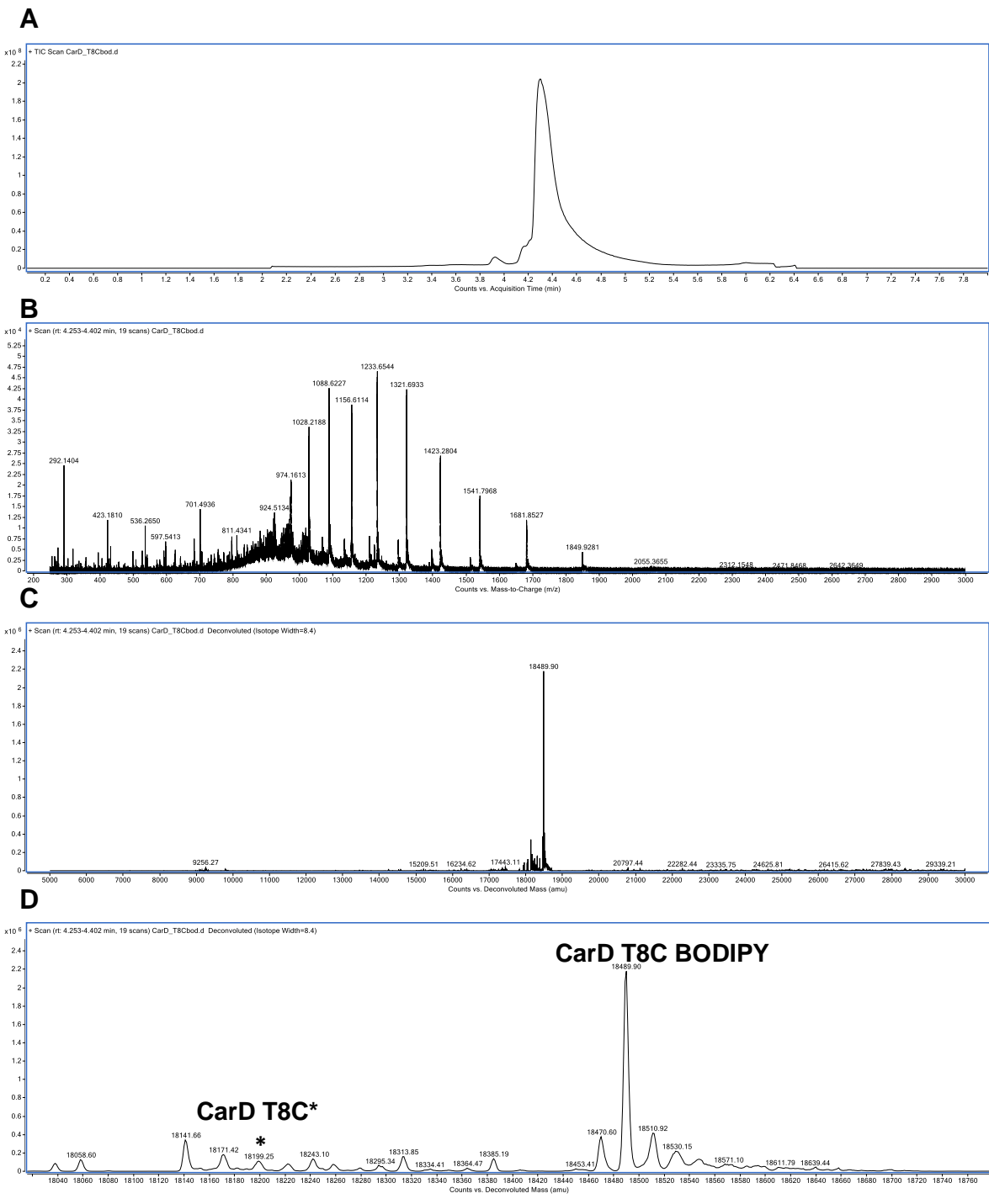
**Appendix Figure IV-A1: LCMS Characterization of WT CarD. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**



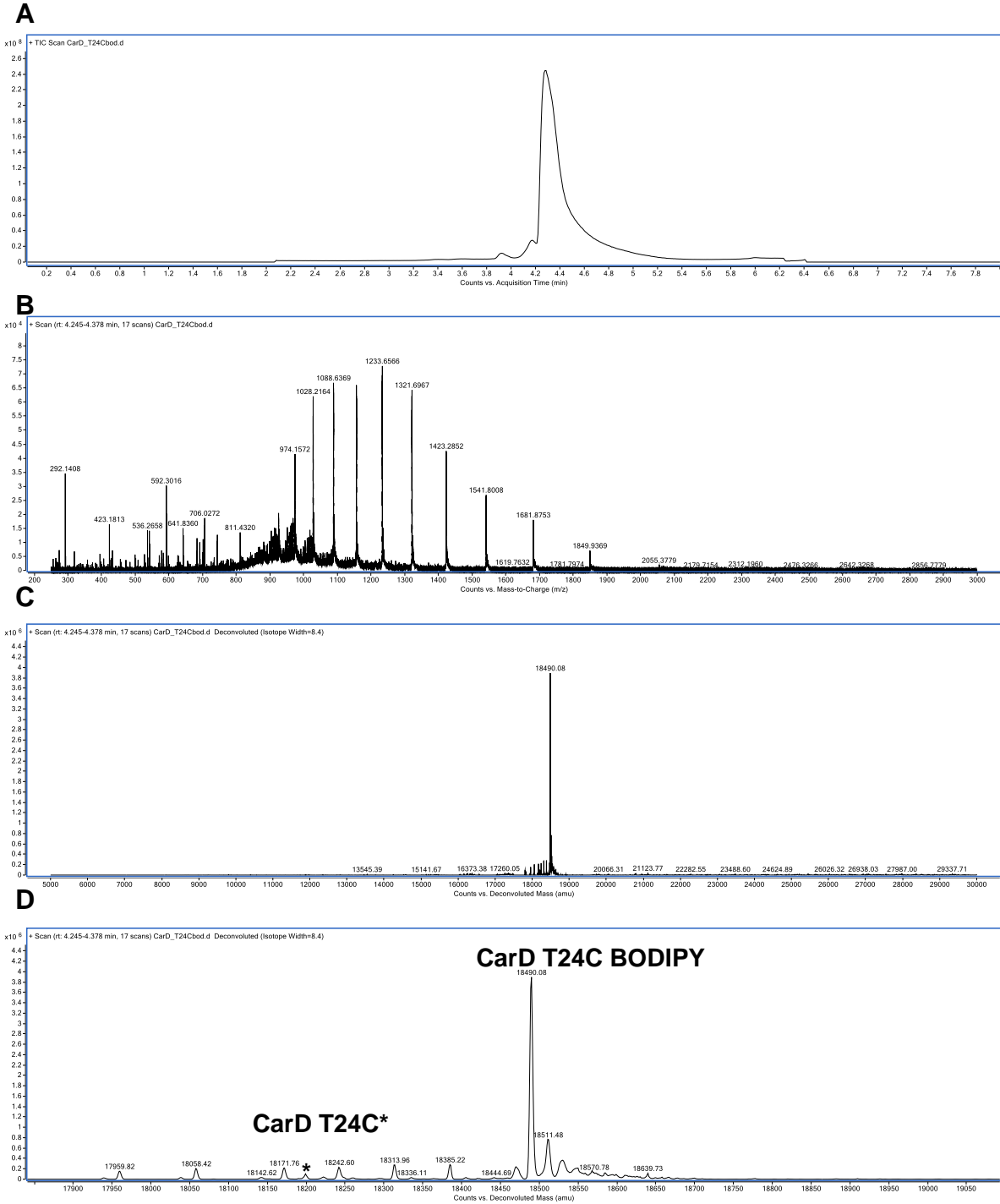
**Appendix Figure IV-A2: LCMS Characterization of CarD R47E. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**



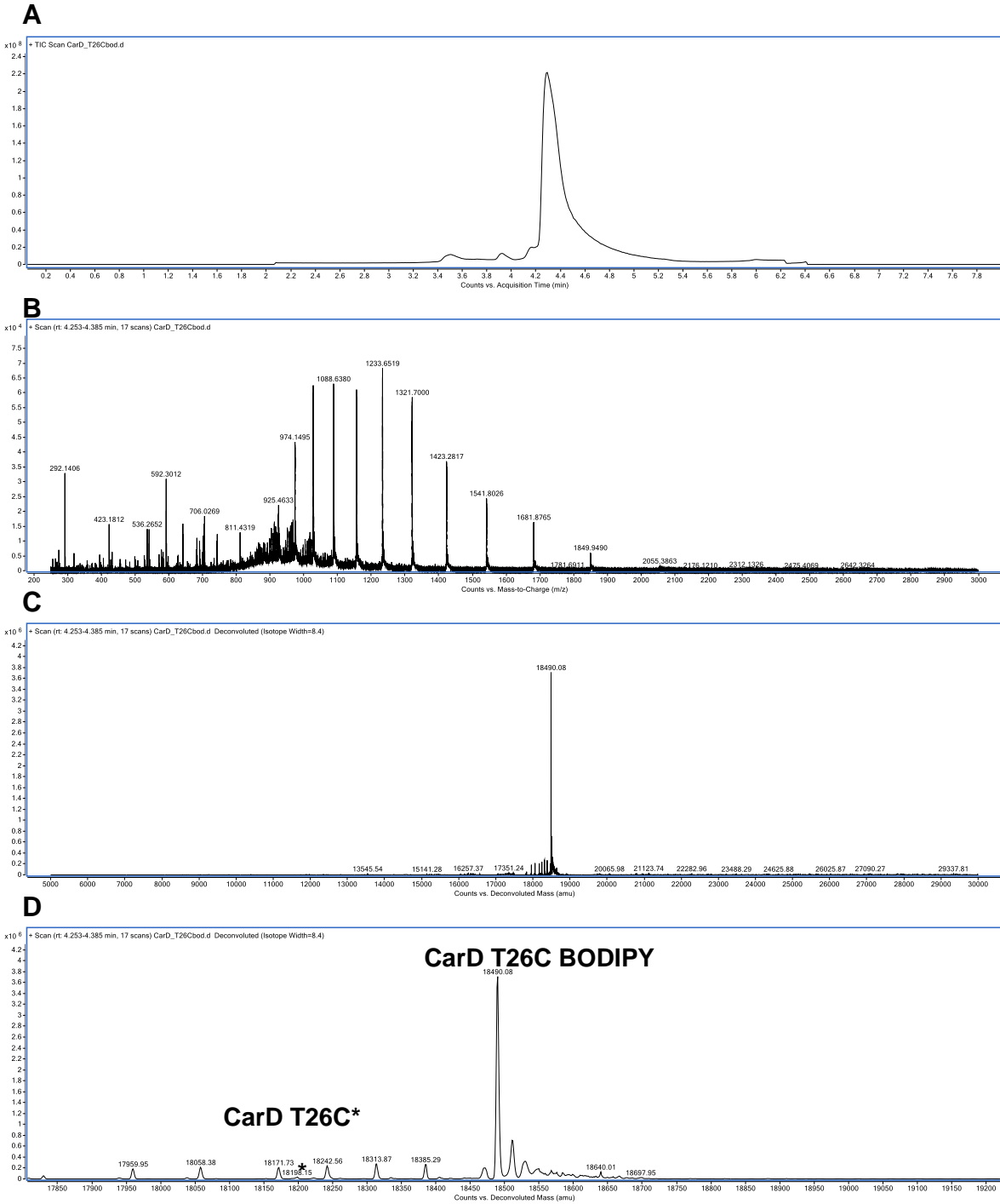
**Appendix Figure IV-A3: LCMS Characterization of CarD K90A. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**



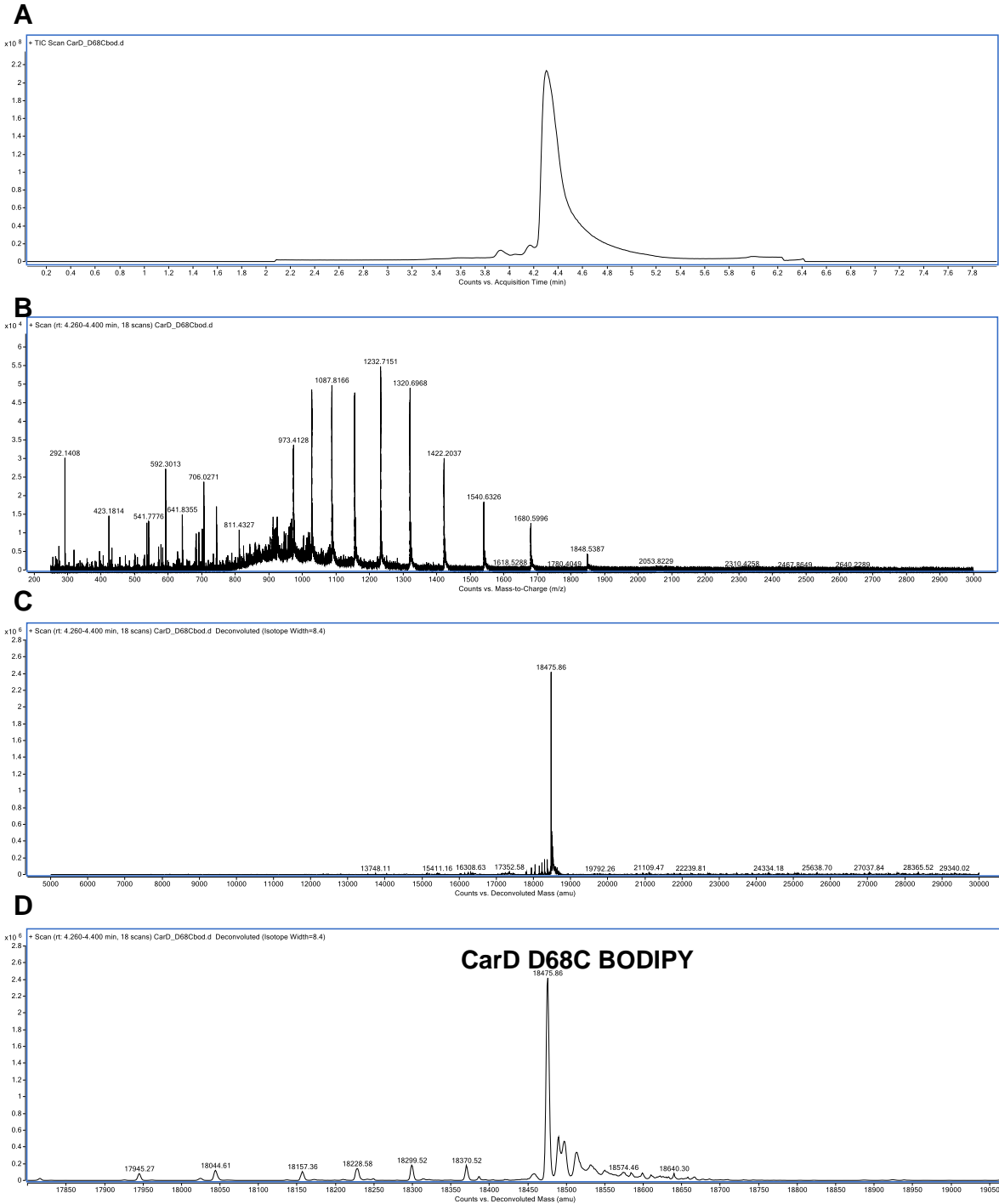
**Appendix Figure IV-A4: LCMS Characterization of CarD T8C BODIPY. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**



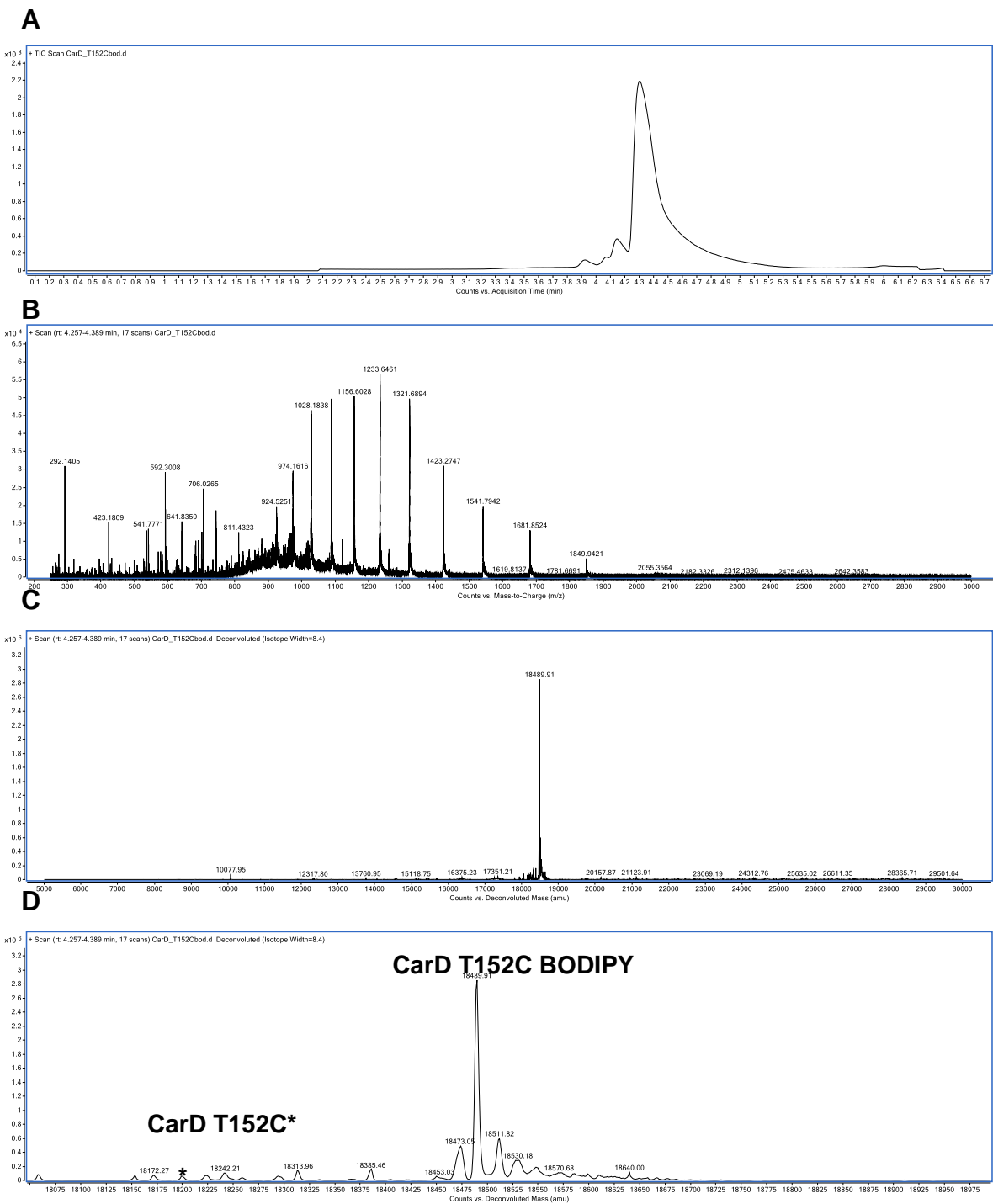
**Appendix Figure IV-A5: LCMS Characterization of CarD T24C BODIPY. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**



**Appendix Figure IV-A6: LCMS Characterization of CarD T26C BODIPY. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**

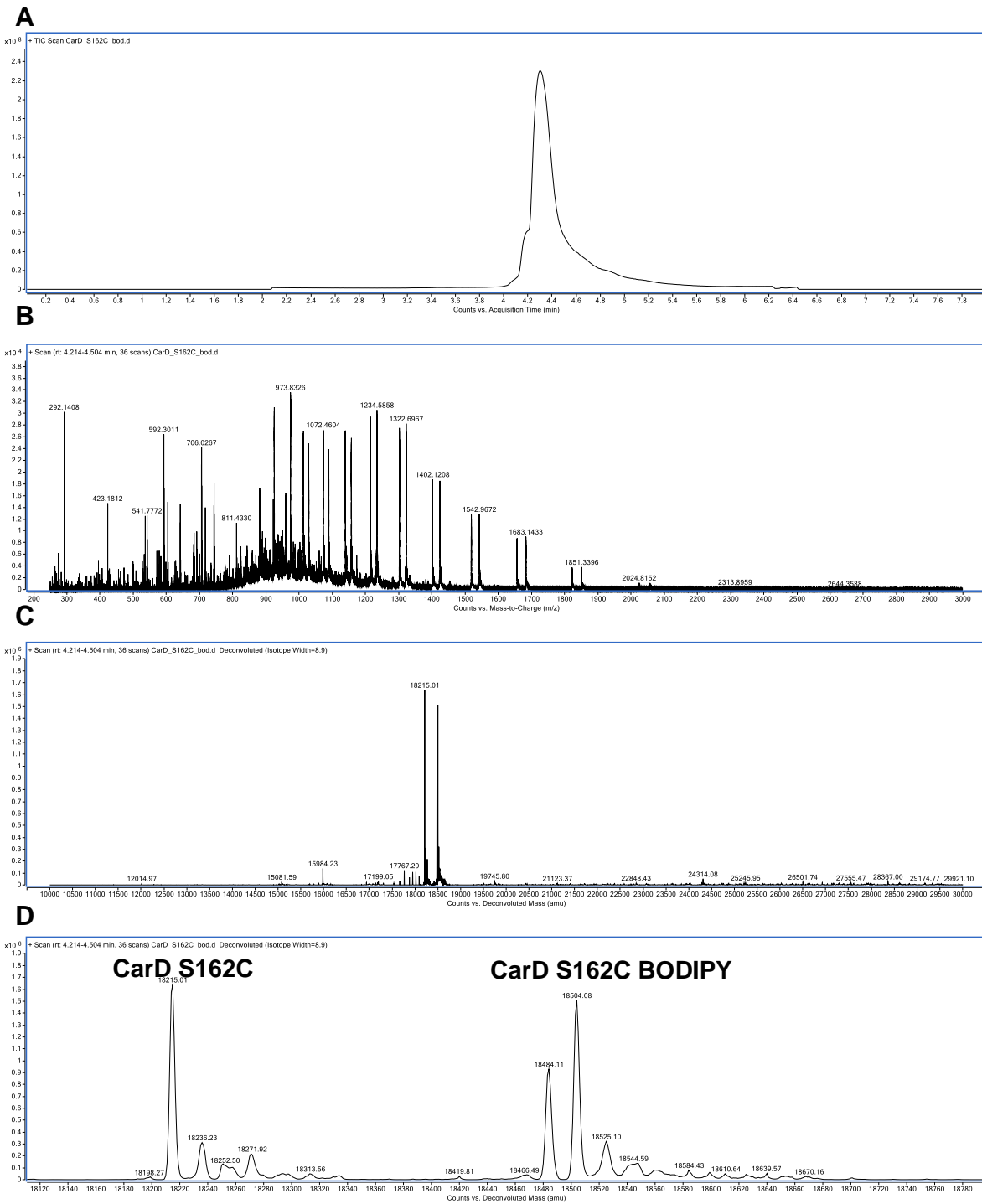


**Appendix Figure IV-A7: LCMS Characterization of CarD D68C BODIPY. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**

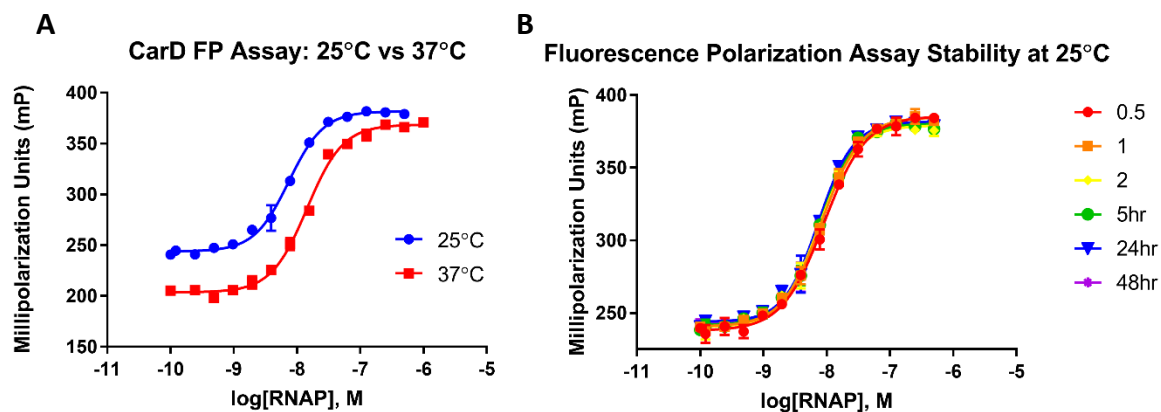


**Appendix Figure IV-A8: LCMS Characterization of CarD T152C BODIPY. A) LC trace, B) fragmentation by MS, C) deconvoluted protein m/z, and D) magnification of peaks proximal to the protein m/z.**

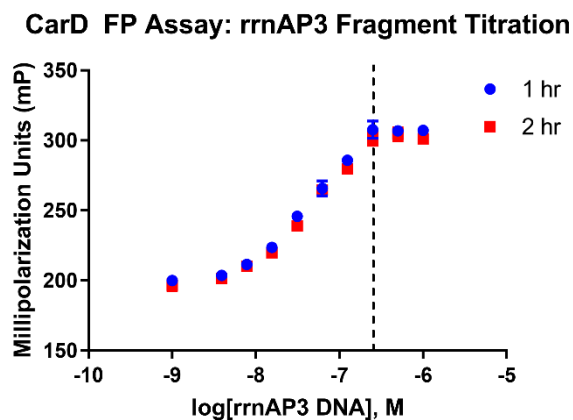




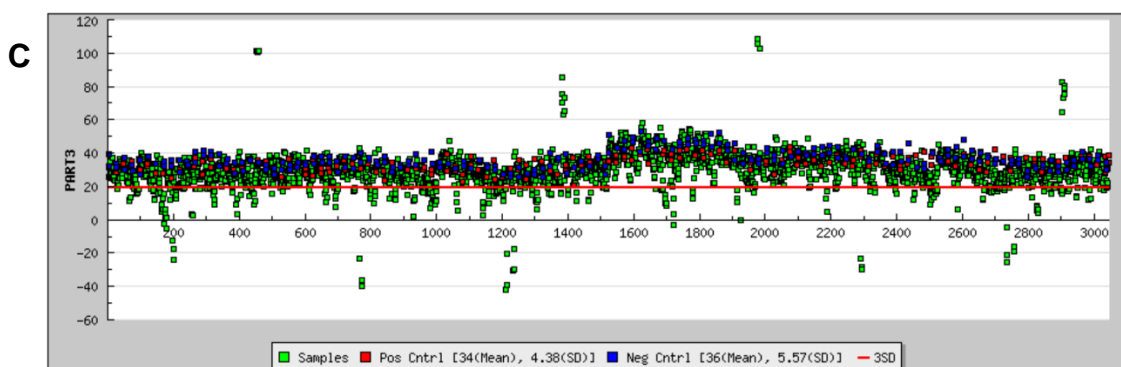
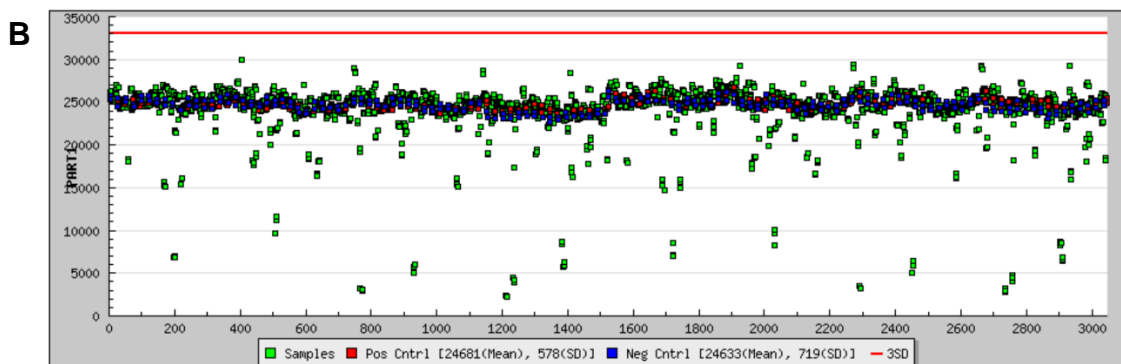
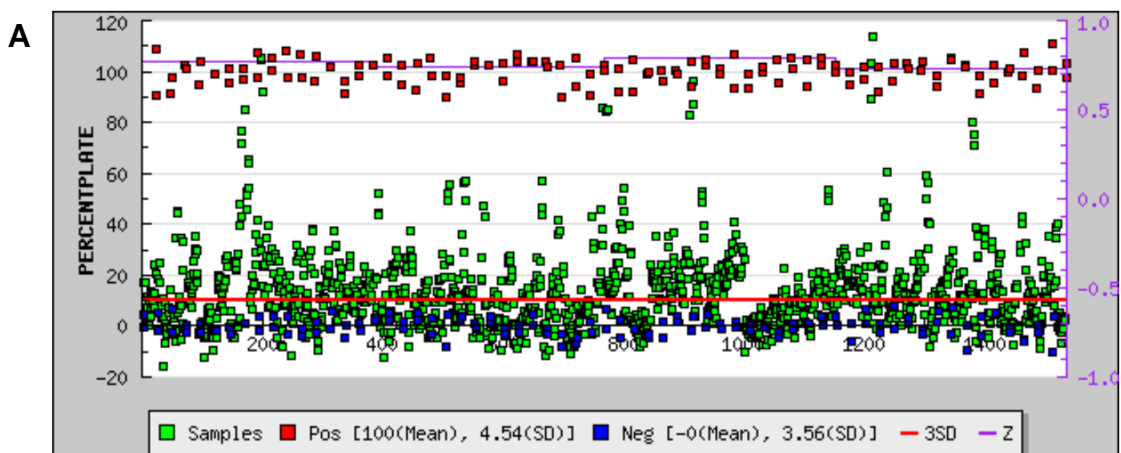
**Appendix Figure IV-A9: LCMS Characterization of CarD S162C BODIPY. A)** LC trace, **B)** fragmentation by MS, **C)** deconvoluted protein m/z, and **D)** magnification of peaks proximal to the protein m/z.



**Appendix Figure IV-A10: CarD FP Assay at 25°C. A)** Effect of temperature on CarD FP Assay **B)** CarD FP assay temporal stability over 48-hours.

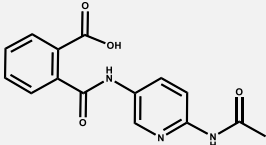
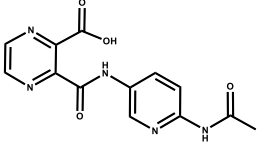
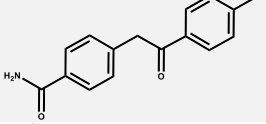
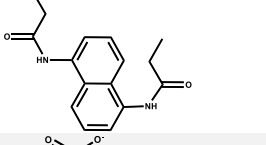
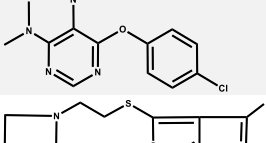
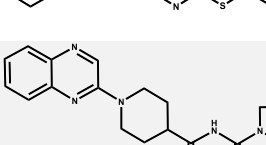
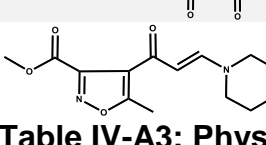



**Appendix Figure IV-A11: Titration of rrnAP3 DNA in CarD FP assay.** RNAP is present at 15 nM and CarD BODIPY at 3 nM. The dashed line is at 250 nM and represents the threshold for the minimum amount of rrnAP3 DNA which needs to be present to maintain signal stability.

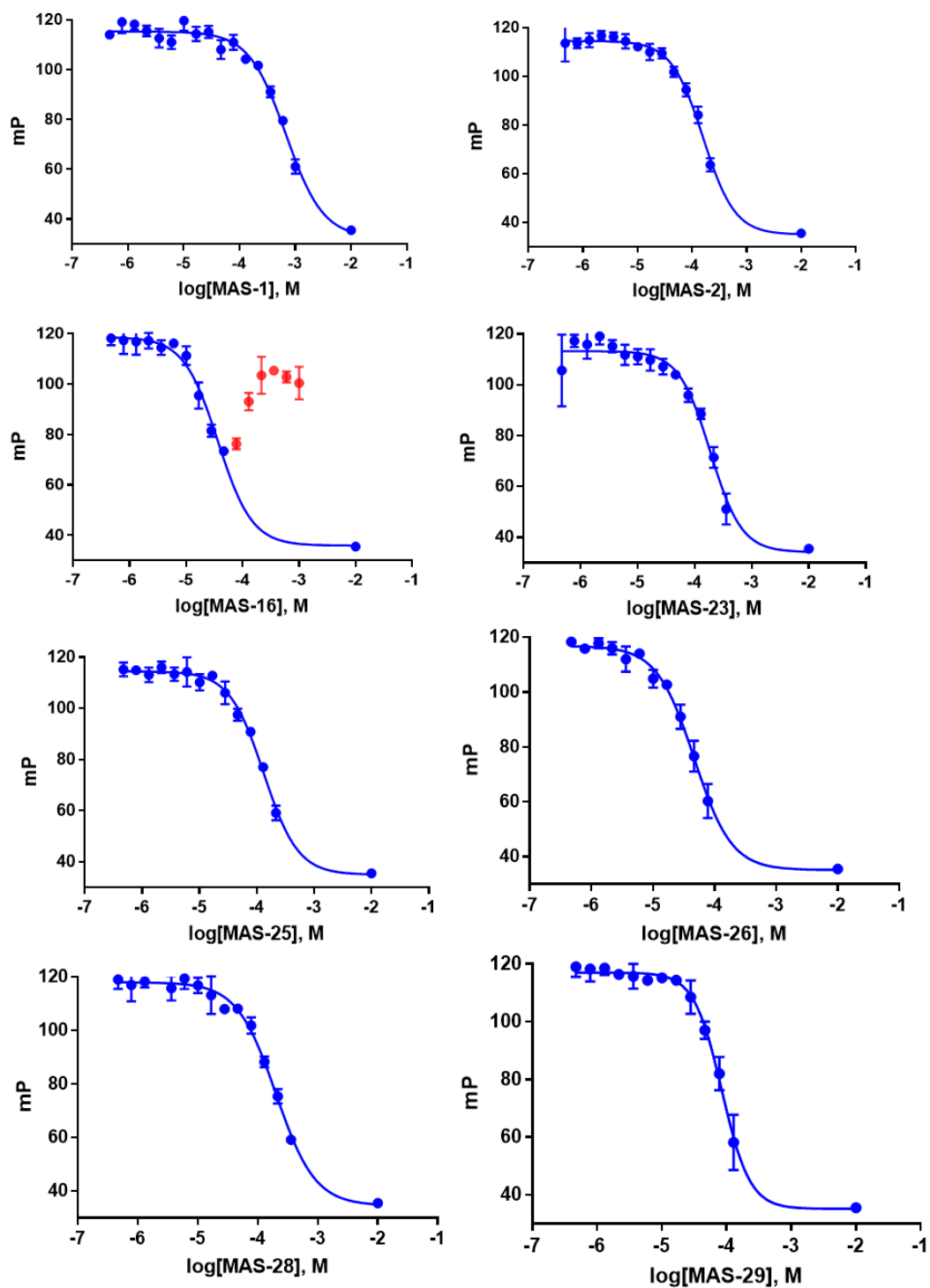


**Appendix Figure IV-A12: Retest plate data for confirmation from MB24K primary.**

**A)** All compounds which were tested for confirmation in triplicate. **B)** Effect of compounds tested against Card BODIPY alone on parallel fluorescence intensity in the counter screen. **C)** Effect of compounds tested against Card BODIPY alone on mP in the counter screen.

Name	Structure	Molecular Weight (g/mol)	cLogP	H-Bond Acceptors	H-Bond donors	TPSA
MAS-1		299.29	1.35	7	3	108.39
MAS-2		301.26	-0.55	9	3	134.17
MAS-16		273.72	2.67	3	1	60.16
MAS-23		270.33	3.17	4	2	58.20
MAS-25		294.70	1.78	7	0	84.07
MAS-26		327.50	2.51	4	1	141.00
MAS-28		353.43	2.07	7	1	78.43
MAS-29		278.31	0.94	6	0	72.64

Appendix Table IV-A3: Physicochemical properties of reconfirmed hits from MB24K HTS.



**Appendix Figure IV-A13: Curves used to generate IC<sub>50</sub> values for hits from the MB24K HTS. Points deleted from MAS-16 due to compound insolubility are indicated in red.**

## CHAPTER V

### Concluding Summary and Future Directions

My dissertation work revolved around understanding the therapeutic potential of transcription in *Mycobacterium tuberculosis*. My first scientific goal was to acquire highly pure and catalytically active MTB RNA polymerase (RNAP), which would then be used in a series of biochemical assays. Using this purified enzyme, my next goal was to characterize the intrinsic properties of MTB RNAP to help explain the clinical prevalence of certain RMP<sup>R</sup> mutations. My final goal was to develop a fluorescence polarization assay to monitor binding of transcriptional regulator CarD to the RPo complex and to use this assay for high throughput therapeutic discovery.

Initial attempts to use a polycistronic expression vector (pMTBRP) containing all the core subunits of RNAP were unsuccessful due to disproportionate overexpression of the subunits.<sup>1</sup> A new methodology which used multiple vectors, some of which contained two of the RNAP subunits, was shown to be effective for overexpression of all subunits; therefore, efforts to optimize the pMTBRP expression system were discontinued (see **Figure II-1**).<sup>2</sup> Several iterations of the multiple vector system were attempted (see **Figure II-2-4**) and a configuration involving three vectors which allowed for enrichment of stoichiometric MTB RNAP holoenzyme was identified (see **Figure II-4**). Initial characterization of the purified enzyme activity with primer extension analysis found that these purified complexes has very low specific activities (see **Figure II-5C**). A zinc-binding

domain (ZBD) in the  $\beta'$ -subunit of RNAP is involved in the proper folding of the subunit and assembly of catalytically-competent RNAP complex.<sup>3</sup> This prompted the supplementation of growth media with ZnSO<sub>4</sub>. Complexes purified under these conditions were significantly more active than those purified without Zn<sup>+2</sup> (see **Figure II-5C**). Purified complexes were also devoid of chromosomally-encoded *E. coli* RNAP subunit contamination (see **Appendix Figure II-5**). Therefore, a method by which highly pure and catalytically active MTB RNAP was successfully developed.

Using this methodology, WT and RMP<sup>R</sup> mutant MTB RNAPs were expressed and purified. A plasmid-based, fluorescence-detected transcription assay (developed by Dr. Nathan Scharf) was used to characterize the potency of rifampin against both WT and RMP<sup>R</sup> mutant RNAPs.<sup>4</sup> Results indicated that all RMP<sup>R</sup> mutations are highly effective at reducing RMP potency (see **Figure II-6**). The  $\beta$ H445Y RMP<sup>R</sup> was totally resistant to RMP, which was corroborated by structural work performed by Dr. Vadim Molodstov and Dr. Katsuhiko Murakami at Penn State University.<sup>5</sup> The  $\beta$ H445Y (or  $\beta$ H526Y in *E. coli*) mutation induces conformational changes in the loops of the RRDR which causes a partial collapse of the RMP binding cleft (see **Figure II-7C, Appendix Figure II-6**). Their structural work also elucidated structural defects associated with the  $\beta$ S450L and  $\beta$ D435V mutants ( $\beta$ S531L and  $\beta$ D516V in *E. coli*). The  $\beta$ S450L mutant causes disorder in Fork loop 2 which results in solvent exposure to RMP, likely reducing its potency (see **Figure II-7E, F**).<sup>5</sup> This in addition to steric induced clashes with leucine at the mutation position represent the structural basis for the reduced potency of RMP for the  $\beta$ S450L mutant. The  $\beta$ D435V mutation significant changes the electrostatic surface of the RRDR pocket, reducing RMP affinity.

RMP<sup>R</sup> mutants with secondary mutations which compensate for a fitness defect associated with the RMP<sup>R</sup> mutants were also investigated to determine the molecular basis for these fitness defects of the most prevalent RMP<sup>R</sup> mutants observed clinically.<sup>6-</sup>

<sup>8</sup> Interestingly, the RMP<sup>R</sup> mutants have distinct mechanistic and kinetic profiles which may be interpreted to explain the prevalence of the mutations in clinical isolates.<sup>8-11</sup> These studies provide a biochemical basis for the evolution of secondary, compensatory mutations observed in the  $\beta'$ -subunit of RNA polymerase (see **Figure I-5**). We conclude that the main contributing factor is maintaining a balance between RNAP elongation rate and termination efficiency. The significant defect in elongation rate seen for the  $\beta$ S450L mutation allows for increased termination efficiency; but likely contributes to the presence of compensatory mutations which mitigate the elongation rate defect. The defect in Fork loop 2 which was observed in the crystal structures discussed above seems to be the cause of the elongation rate defect in the  $\beta$ S450L (see **Figure II-7E, F, Figure III-2**). Additionally, both the RMP<sup>R</sup> and compensatory mutations lie on opposite sides of the DNA•RNA heteroduplex within the main channel and are likely affecting both elongation and termination by altering the efficiency of the clamp domain which normally stabilizes the RNA•DNA template heteroduplex on RNAP (see **Figure III-5**).<sup>12-13</sup> Mutations in RRDR have been shown to effect RNA slippage, consistent with the heteroduplex•RNAP complex being de-stabilized by their presence.<sup>12, 14</sup> This is also corroborated by our RNA primer hydrolysis data, which shows the  $\beta'$ V483G mutant, which is in the DPBB adjacent to the clamp, causes increased hydrolysis of RNA primer, likely by effecting the DNA•RNA heteroduplex complex stability on RNAP.



Another factor which seems to drive secondary mutations in the  $\beta'$ -subunit is their effect on the open-promoter complex (RPO) stability. All of the mutations are on the same amino acid chain that leads into Fork loop 2. It seems likely that this is contributing to RPO destabilization as Fork loop 2 lies at the downstream fork of the transcription bubble, potentially weakening stability of the RPO at this position.<sup>5</sup> Further studies which probe the structural consequences of the RMP<sup>R</sup> mutant fitness defect compensation by compensatory mutations in the  $\beta'$ -subunit are needed to confirm how the secondary mutations are altering RNAP structure to elucidate the exact molecular mechanism of action of these mutations.

RNAP is a validated antibacterial target; however, traditionally the focus has been on targeting transcription via direct inhibition of RNAP activity (see **Figure I-7**). There are several existing inhibitors which do not directly bind to RNAP. Bicyclomycin is an inhibitor of Rho dependent termination and there are several AraC-family inhibitors.<sup>15-19</sup> AraC acts by recruiting RNAP to a promoter and has been proposed to stimulate RPO formation, though this is contentious as structural work supports a “recruitment only” mechanism through interactions with the C-terminal domain of the RNAP  $\alpha$ -subunit.<sup>20-21</sup> Therefore, regulation of the RNAP open-complex (RPO) formation appears to be an untapped target for novel antibacterial targets. CarD a transcriptional regulator in MTB which stabilizes the RPO through bipartite interactions with both RNAP and promoter DNA (see **Figure I-8, Figure III-1**).<sup>22-23</sup> CarD has been shown to be required for MTB viability and it represents a potentially novel therapeutic target.<sup>24-25</sup>

A fluorescence polarization assay which probes the interaction between CarD and RPO was developed, optimized, and validated (see **Figure IV-1-6**). Due to the differential

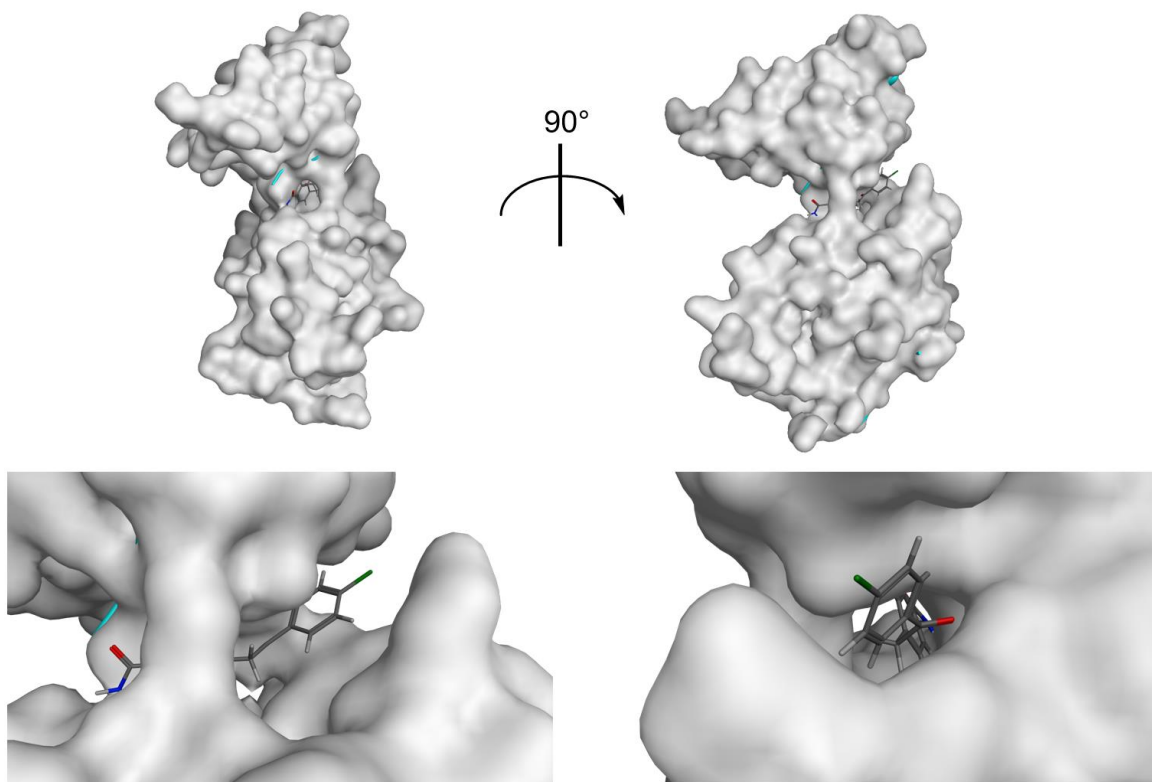
affinity of CarD for free RNAP and the RPo, this assay can be used to not only inhibitors of CarD binding to RPo\*DNA, but also DNA association with RNAP. This is evidenced by the observed inhibition of CarD FP by fidaxomicin which blocks DNA association preventing RNAP clamp closure and RPo formation (see **Figure IV-5**).<sup>26-27</sup>

The CarD FP assay is optimal for HTS. The CarD•RPo FP signal is stable over hours and the assay is robust for HTS. Competition experiments with unlabeled CarD indicate that the fluorophore labeling does not affect CarD binding to RPo. In pilot screen of 237 small molecules demonstrated that this assay has an excellent z-score (0.77, see **Figure IV-6**), though no inhibitors were identified from this library. A larger library, MB24K (see **Figure IV-7**), which consists of 23,320 compounds selected by medicinal chemists, was screened. The assay had an excellent overall z-score (0.73, see **Figure IV-7**); however, there was significant spectral overlap between our fluorescent probe (BODIPY FL) and compounds in the library which lead to an unprocessed hit rate of 4.26%.

Triage of compounds to remove fluorescence-interfering compounds using the parallel fluorescence intensity, as well as, those which were identified by NIH flags brought our hit rate from the primary screen down to 1.8% which is still higher than the anticipated 1% (see **Figure IV-7B, 8**). Though the hit rate was high we decided to move forward with this set and eliminate compounds during confirmation and in the counter screen. Roughly 50% of compounds tested during confirmation did not reconfirm bringing our hit rate down to 0.9% (see **Appendix Figure IV-12A, B**). Again, we had to eliminate several compounds because of potential interference issues and also PAIN compounds (see **Figure IV-8**). A counter screen was conducted to identify false positives due to CarD-Bodipy interference and by the end of this process we were left with 109 out of 23,320

compounds or a 0.47% triaged hit rate (see **Appendix Figure IV-12C**). Of these, 100 compounds were tested for concentration response (CRC) and we had 59 actives. We ordered 33 fresh powders for reconfirmation.

Of the 33 powders we ordered, only 8 reconfirmed in our FP assay, and these are relatively weak inhibitors (see **Table IV-1**). However, protein-protein interactions are difficult to target because they usually have large and flat interfaces which are difficult for small molecules to bind with any significant affinity. Therefore, it seems fortunate to have found anything in this limited screen. MAS-16 appears to selectively bind to CarD (see **Figure IV-9**) in a direct binding SPR assay which is promising. Initial attempts at modeling MAS-16 in the CarD structure show that MAS-16 can bind in the region between the N-terminal RNAP interacting domain and the C-terminal DNA binding domain. This potential binding mode is not yet confirmed; however, in addition to validating the model, this could serve as a preliminary model identify analogs of MAS-16 which may have increased potency. Several other biochemical assays should be performed on the remaining hits to determine whether they act by preventing RNAP association with promoter DNA, which is another potential mechanism of action which can be monitored in the FP assay (see **Figure IV-5**).



**Figure V-1: Model of MAS-16 docked into MTB CarD using MOE. A)** Full length MTB CarD with MAS-16 virtually docked (PDB: 4KBM). **B)** Close up of proposed binding site at the interface of the RNAP and DNA binding domains.

In conclusion I have prepared several biochemical assays and their constituents which have allowed us to probe MTB RNAP mechanistically. We have strong biochemical evidence which explains the observed clinical prevalence of particular RMP<sup>R</sup> mutants and have provided strong evidence supporting a rationale for the evolution of fitness compensating mutations observed in RMP<sup>R</sup> MTB. I have also set the foundation for the discovery of novel inhibitors of the MTB transcriptional regulator CarD. This work is still in the initial discovery phase and will be the foundation for a campaign to try and identify more selective and potent inhibitors of CarD.

## References

1. Gill, S. K.; Garcia, G. A., Rifamycin inhibition of WT and Rif-resistant Mycobacterium tuberculosis and Escherichia coli RNA polymerases in vitro. *Tuberculosis (Edinb)* **2011**, *91* (5), 361-9.
2. Banerjee, R.; Rudra, P.; Prajapati, R. K.; Sengupta, S.; Mukhopadhyay, J., Optimization of recombinant Mycobacterium tuberculosis RNA polymerase expression and purification. *Tuberculosis (Edinb)* **2014**, *94* (4), 397-404.
3. Markov, D.; Naryshkina, T.; Mustaev, A.; Severinov, K., A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes Dev* **1999**, *13* (18), 2439-48.
4. Scharf, N. T.; Molodtsov, V.; Kontos, A.; Murakami, K. S.; Garcia, G. A., Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase Discovered from High-Throughput Screening. *SLAS Discov* **2017**, *22* (3), 287-297.
5. Molodtsov, V.; Scharf, N. T.; Stefan, M. A.; Garcia, G. A.; Murakami, K. S., Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis. *Mol Microbiol* **2017**, *103* (6), 1034-1045.
6. Song, T.; Park, Y.; Shamputa, I. C.; Seo, S.; Lee, S. Y.; Jeon, H. S.; Choi, H.; Lee, M.; Glynn, R. J.; Barnes, S. W.; Walker, J. R.; Batalov, S.; Yusim, K.; Feng, S.; Tung, C. S.; Theiler, J.; Via, L. E.; Boshoff, H. I.; Murakami, K. S.; Korber, B.; Barry, C. E., 3rd; Cho, S. N., Fitness costs of rifampicin resistance in Mycobacterium tuberculosis are amplified under conditions of nutrient starvation and compensated by mutation in the beta' subunit of RNA polymerase. *Mol Microbiol* **2014**, *91* (6), 1106-19.
7. Brandis, G.; Hughes, D., Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *J Antimicrob Chemother* **2013**, *68* (11), 2493-7.

8. Comas, I.; Borrell, S.; Roetzer, A.; Rose, G.; Malla, B.; Kato-Maeda, M.; Galagan, J.; Niemann, S.; Gagneux, S., Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* **2011**, *44* (1), 106-10.
9. Tang, K.; Sun, H.; Zhao, Y.; Guo, J.; Zhang, C.; Feng, Q.; He, Y.; Luo, M.; Li, Y.; Sun, Q., Characterization of rifampin-resistant isolates of *Mycobacterium tuberculosis* from Sichuan in China. *Tuberculosis (Edinb)* **2013**, *93* (1), 89-95.
10. Rahmo, A.; Hamdar, Z.; Kasaa, I.; Dabboussi, F.; Hamze, M., Genotypic detection of rifampicin-resistant *M. tuberculosis* strains in Syrian and Lebanese patients. *J Infect Public Health* **2012**, *5* (6), 381-7.
11. Poudel, A.; Nakajima, C.; Fukushima, Y.; Suzuki, H.; Pandey, B. D.; Maharjan, B.; Suzuki, Y., Molecular characterization of multidrug-resistant *Mycobacterium tuberculosis* isolated in Nepal. *Antimicrob Agents Chemother* **2012**, *56* (6), 2831-6.
12. Zhou, Y. N.; Lubkowska, L.; Hui, M.; Court, C.; Chen, S.; Court, D. L.; Strathern, J.; Jin, D. J.; Kashlev, M., Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem* **2013**, *288* (4), 2700-10.
13. Kireeva, M. L.; Domecq, C.; Coulombe, B.; Burton, Z. F.; Kashlev, M., Interaction of RNA polymerase II fork loop 2 with downstream non-template DNA regulates transcription elongation. *J Biol Chem* **2011**, *286* (35), 30898-910.
14. Strathern, J.; Malagon, F.; Irvin, J.; Gotte, D.; Shafer, B.; Kireeva, M.; Lubkowska, L.; Jin, D. J.; Kashlev, M., The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in *S. cerevisiae*. *J Biol Chem* **2013**, *288* (4), 2689-99.
15. Skordalakes, E.; Brogan, A. P.; Park, B. S.; Kohn, H.; Berger, J. M., Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure* **2005**, *13* (1), 99-109.
16. Woodbrey, A. K.; Onyango, E. O.; Pellegrini, M.; Kovacicova, G.; Taylor, R. K.; Gribble, G. W.; Kull, F. J., A new class of inhibitors of the AraC family virulence regulator *Vibrio cholerae* ToxT. *Sci Rep* **2017**, *7*, 45011.

17. Yang, J.; Hocking, D. M.; Cheng, C.; Dogovski, C.; Perugini, M. A.; Holien, J. K.; Parker, M. W.; Hartland, E. L.; Tauschek, M.; Robins-Browne, R. M., Disarming bacterial virulence through chemical inhibition of the DNA binding domain of an AraC-like transcriptional activator protein. *J Biol Chem* **2013**, *288* (43), 31115-26.
18. Koppolu, V.; Osaka, I.; Skredenske, J. M.; Kettle, B.; Hefty, P. S.; Li, J.; Egan, S. M., Small-molecule inhibitor of the *Shigella flexneri* master virulence regulator VirF. *Infect Immun* **2013**, *81* (11), 4220-31.
19. Emanuele, A. A.; Garcia, G. A., Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the *Shigella flexneri* Virulence Regulator VirF. *PLoS One* **2015**, *10* (9), e0137410.
20. Johnson, C. M.; Schleif, R. F., Cooperative action of the catabolite activator protein and AraC in vitro at the araFGH promoter. *J Bacteriol* **2000**, *182* (7), 1995-2000.
21. Benoff, B.; Yang, H.; Lawson, C. L.; Parkinson, G.; Liu, J.; Blatter, E.; Ebright, Y. W.; Berman, H. M.; Ebright, R. H., Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* **2002**, *297* (5586), 1562-6.
22. Bae, B.; Chen, J.; Davis, E.; Leon, K.; Darst, S. A.; Campbell, E. A., CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* **2015**, *4*.
23. Gulten, G.; Sacchettini, J. C., Structure of the Mtb CarD/RNAP beta-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure* **2013**, *21* (10), 1859-69.
24. Weiss, L. A.; Harrison, P. G.; Nickels, B. E.; Glickman, M. S.; Campbell, E. A.; Darst, S. A.; Stallings, C. L., Interaction of CarD with RNA polymerase mediates *Mycobacterium tuberculosis* viability, rifampin resistance, and pathogenesis. *J Bacteriol* **2012**, *194* (20), 5621-31.
25. Garner, A. L.; Rammohan, J.; Huynh, J. P.; Onder, L. M.; Chen, J.; Bae, B.; Jensen, D.; Weiss, L. A.; Manzano, A. R.; Darst, S. A.; Campbell, E. A.; Nickels, B. E.; Galburt, E. A.; Stallings, C. L., Effects of Increasing the Affinity of CarD for RNA Polymerase on *Mycobacterium tuberculosis* Growth, rRNA Transcription, and Virulence. *J Bacteriol* **2017**, *199* (4).

26. Boyaci, H.; Chen, J.; Lilic, M.; Palka, M.; Mooney, R. A.; Landick, R.; Darst, S. A.; Campbell, E. A., Fidaxomicin jams Mycobacterium tuberculosis RNA polymerase motions needed for initiation via RbpA contacts. *Elife* **2018**, *7*.
27. Lin, W.; Das, K.; Degen, D.; Mazumder, A.; Duchi, D.; Wang, D.; Ebright, Y. W.; Ebright, R. Y.; Sineva, E.; Gigliotti, M.; Srivastava, A.; Mandal, S.; Jiang, Y.; Liu, Y.; Yin, R.; Zhang, Z.; Eng, E. T.; Thomas, D.; Donadio, S.; Zhang, H.; Zhang, C.; Kapanidis, A. N.; Ebright, R. H., Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Mol Cell* **2018**, *70* (1), 60-71 e15.